**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, C07K 14/47, A01K 67/027, C07K 16/18, C12Q 1/68, G06F 17/30

A1

(11) International Publication Number: WO 00/63375

(43) International Publication Date: 26 October 2000 (26.10.00)

(21) International Application Number: PCT/IB00/00562

(22) International Filing Date: 20 April 2000 (20.04.00)

(30) Priority Data:
60/130,217 20 April 1999 (20.04.99) US

(71) Applicant *(for all designated States except US)*: GENSET [FR/FR]; Intellectual Property Dept., 24, rue Royale, F–75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F–92170 Vanves (FR). DUFAURE–GARE, Isabelle [FR/FR]; 1 bis, rue d'Alesia, F–75014 Paris (FR). GREL, Pascale [FR/FR]; 45, boulevard Alsace–Lorraine, F–91170 Viry–Chatillon (FR).

(74) Common Representative: GENSET; Intellectual Property Dept., 24, rue Royale, F–75008 Paris (FR).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: DNA ENCODING A KINESIN–LIKE PROTEIN (HKLP) COMPRISING BIALLELIC MARKERS

(57) Abstract

The present invention is directed to polynucleotides encoding a human kinesin–like polypeptide as well as a regulatory region located at the 3'–end of said coding region. The invention also concerns polypeptides encoded by the kinesin–like gene. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the *HKLP* gene useful in genetic analysis.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# DNA ENCODING A KINESIN-LIKE PROTEIN (HKLP) COMPRISING BIALLELIC MARKERS

## FIELD OF THE INVENTION

The present invention is directed to polynucleotides encoding a human kinesin-like polypeptide as well as a regulatory region located at the 3'-end of said coding region. The invention also concerns polypeptides encoded by the kinesin-like gene. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the *HKLP* gene useful in genetic analysis.

## BACKGROUND OF THE INVENTION

The kinesins are mechanochemical proteins utilizing chemical energy from ATP hydrolysis to generate mechanical force. The kinesins can bind to and move on microtubules in the presence of ATP. The ability to move on microtubules has led to the classification of kinesins as microtubule motor proteins. The kinesins play important roles in intracellular transport and cell division.

Several kinesin proteins are involved in vesicle/organelle transport in neurons, and mutants of kinesin in *Drosphila* show impaired neuronal function. In humans, defects in kinesin-encoding genes could cause neurological disorders or syndromes of clinical importance.

The kinesin proteins carry out or facilitate movements of the chromosomes and spindle in meiosis and mitosis. Defective meiotic kinesins in humans may be the causes of infertility, spontaneous abortion, neonatal chromosome disorders, and aneuploidy. In mitotically dividing cells, mutations in kinesin proteins could cause somatic abnormalities or cellular transformation, including neoplasia.

Finally, the kinesins could be involved in developmental processes as the localization of some morphogens has been shown to be microtubule-dependent.

The KIF kinesin superfamily proteins have been identified as candidate motor proteins involved in organelle transport.

Among the KIFs, the murine KIF1A protein has been proposed as a transporter of synaptic vesicle precursors. KIF1A disruption assays in mice allowed to show that KIF1A is involved in the transport of a synaptic vesicle precursor and that KIF1A-mediated axonal transport plays a critical role in viability, maintenance, and function of neurons, particularly mature neurons (Yonekawa et al., 1998). The murine KIF1B protein is co-localized with mitochondria *in vivo* and could be involved in the transport of mitochondria (Nangaku et al., 1994).

## SUMMARY OF THE INVENTION

The present invention pertains to nucleic acid molecules comprising the genomic sequence of a novel human gene which encodes a kinesin-like protein and which has been named *HKLP* by the inventors. The *HKLP* presents homology with murine *KIF1A* and *KIF1B*. The *HKLP* genomic

sequence comprises regulatory sequence located downstream (3'-end) of the transcribed portion of said gene, these regulatory sequences being also part of the invention.

The invention also deals with the complete cDNA sequence encoding the HKLP protein, as well as with the corresponding translation product.

Oligonucleotide probes or primers hybridizing specifically with a *HKLP* genomic or cDNA sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

A further object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described above, and in particular of recombinant vectors comprising a *HKLP* regulatory sequence or a sequence encoding a HKLP protein, as well as of cell hosts and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

The invention is also directed to biallelic markers that are located within the *HKLP* genomic sequence or that are in linkage disequilibrium with the *HKLP* gene, these biallelic markers representing useful tools in order to identify a statistically significant association between specific alleles of *HKLP* gene and diseases, for example cancer and neurological disorders. These association methods are within the scope of the invention.

Finally, the invention is directed to methods for the screening of substances or molecules that inhibit the expression of *HKLP*, as well as with methods for the screening of substances or molecules that interact with a HKLP polypeptide or that modulate the activity of a HKLP polypeptide.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of an exemplary computer system.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

## BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING

SEQ ID Nos 1 and 2 contain the genomic sequence of the *HKLP* gene comprising the exons and introns, and the 3' regulatory region (downstream untranscribed region).

SEQ ID No 3 contains a cDNA sequence of the *HKLP* gene.

SEQ ID No 4 contains the amino acids sequence encoding by the cDNA of SEQ ID No 3.

SEQ ID Nos 5, 6, 7 and 8 respectively contain the nucleotide sequence of the amplicons 10-265, 10-266, 12-592 and 12-783.

SEQ ID No 9 contains a primer containing the additional PU 5' sequence described further in Example 2.

SEQ ID No 10 contains a primer containing the additional RP 5' sequence described further in Example 2.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an thymine.

The nucleotide code of the original allele for each biallelic marker is the following table:

| Biallelic marker | Original allele |
|---|---|
| 12-809-119 | C |
| 12-805-115 | A |
| 12-790-396 | G |
| 12-791-211 | G |
| 12-803-125 | T |
| 99-33040-321 | T |
| 12-810-77 | A |
| 12-787-103 | A |
| 12-793-383 | T |
| 12-792-233 | A |
| 99-41009-244 | A |
| 99-41009-111 | C |
| 12-593-174 | T |
| 12-589-152 | T |
| 12-785-200 | T |
| 12-785-393 | A |
| 12-588-103 | G |
| 12-603-191 | T |
| 12-586-414 | G |
| 12-602-196 | T |
| 12-602-350 | C |
| 12-587-379 | A |
| 12-596-124 | G |
| 12-808-52 | A |
| 12-808-75 | G |

## DETAILED DESCRIPTION OF THE INVENTION

The present invention concerns polynucleotides and polypeptides related to the *HKLP* gene. Oligonucleotide probes and primers hybridizing specifically with a genomic or a cDNA sequence of *HKLP* are also part of the invention. A further object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described in the present invention, and in particular recombinant vectors comprising a regulatory region of *HKLP* or a sequence encoding the HKLP protein, as well as cell hosts comprising said nucleic acid sequences or recombinant vectors. The invention also encompasses methods of screening of molecules which inhibit the expression of the *HKLP* gene or which modulate the activity of the HKLP protein. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents.

The invention also concerns *HKLP*-related biallelic markers which can be used in any method of genetic analysis including linkage studies in families, linkage disequilibrium studies in populations and association studies of case-control populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in complex traits.

### Definitions

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

The terms "*HKLP* gene", when used herein, encompasses genomic, mRNA and cDNA sequences encoding the *HKLP* protein, including the untranslated regulatory regions of the genomic DNA.

The term "heterologous protein", when used herein, is intended to designate any protein or polypeptide other than the HKLP protein. More particularly, the heterologous protein is a compound which can be used as a marker in further experiments with a *HKLP* regulatory region.

The term "isolated" requires that the material be removed from its original environment (e. g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude. The term "purified" is used herein to describe a polynucleotide or

polynucleotide vector of the invention which has been separated from other compounds including, but not limited to other nucleic acids, carbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

The term "purified" is used herein to describe a polypeptide of the invention which has been separated from other compounds including, but not limited to nucleic acids, lipids, carbohydrates and other proteins. A polypeptide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure polypeptide typically comprises about 50%, preferably 60 to 90% weight/weight of a protein sample, more usually about 95%, and preferably is over about 99% pure. Polypeptide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to

refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human".

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', $F(ab)_2$, and $F(ab')_2$ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a HKLP polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by Geysen et al. 1984; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

As used interchangeably herein, the terms "nucleic acids", "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

A sequence which is "operably linked" to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to a disease, a beneficial response to or side effects related to a treatment. Preferably, said trait can be, without to be limited to, cancers, developmental diseases, and neurological diseases.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms may be homozygous or heterozygous for an allelic form.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population which are heterozygous at a particular allele. In a biallelic system, the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where $P_a$ is the frequency of the least common allele. In order to be useful in genetic studies, a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention, a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention, a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which may be associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide also gives rise to single nucleotide polymorphisms. In the context of the present invention, "single nucleotide polymorphism"

preferably refers to a single nucleotide substitution. Typically, between different individuals, the polymorphic site may be occupied by two different nucleotides.

The term "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a single nucleotide polymorphism having two alleles at a fairly high frequency in the population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Typically, the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker".

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym

from "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

**Variants and Fragments**

1- Polynucleotides

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of a *HKLP* gene containing one or more biallelic markers according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences which are at least 95% identical to a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-3 or to any polynucleotide fragment of at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-3, and preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-3 or to any polynucleotide fragment of at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID No 1-3.

Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature HKLP protein, or those in which the polynucleotides encode polypeptides which maintain or increase a particular biological activity, while reducing a second biological activity

A polynucleotide fragment is a polynucleotide having a sequence that is entirely the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of a *HKLP* gene, and variants thereof. The fragment can be a portion of an intron or an exon of a *HKLP* gene. It can also be a portion of the regulatory regions of *HKLP*. Preferably, such fragments comprise at least one of the biallelic markers A1 to A32 or the complements thereto or a biallelic marker in linkage disequilibrium therewith.

Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. Indeed, several of these fragments may be present within a single larger polynucleotide.

Optionally, such fragments may consist of, or consist essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length.

2- Polypeptides

The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein, including mutated HKLP proteins.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated HKLP is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated HKLP, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated HKLP or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

A polypeptide fragment is a polypeptide having a sequence that entirely is the same as part but not all of a given polypeptide sequence, preferably a polypeptide encoded by a *HKLP* gene and variants thereof.

In the case of an amino acid substitution in the amino acid sequence of a polypeptide according to the invention, one or several amino acids can be replaced by "equivalent" amino acids. The expression "equivalent" amino acid is used herein to designate any amino acid that may be substituted for one of the amino acids having similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Generally, the following groups of amino acids represent equivalent changes: (1) Ala, Pro, Gly, Glu, Asp, Gln, Asn, Ser, Thr; (2) Cys, Ser, Tyr, Thr; (3) Val, Ile, Leu, Met, Ala, Phe; (4) Lys, Arg, His; (5) Phe, Tyr, Trp, His.

A specific embodiment of a modified HKLP peptide molecule of interest according to the present invention, includes, but is not limited to, a peptide molecule which is resistant to proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH2NH) reduced

bond, a (NHCO) retro inverso bond, a (CH2-O) methylene-oxy bond, a (CH2-S) thiomethylene bond, a (CH2CH2) carba bond, a (CO-CH2) cetomethylene bond, a (CHOH-CH2) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond. The invention also encompasses a human HKLP polypeptide or a fragment or a variant thereof in which at least one peptide bond has been modified as described above.

Such fragments may be "free-standing", i.e. not part of or fused to other polypeptides, or they may be comprised within a single larger polypeptide of which they form a part or region. However, several fragments may be comprised within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids.

**Identity Between Nucleic Acids Or Polypeptides**

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988; Altschul et al., 1990; Thompson et al., 1994; Higgins et al., 1996; Altschul et al., 1990; Altschul et al., 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul et al., 1990, 1993, 1997). In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992; Henikoff and Henikoff, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990).

**Stringent Hybridization Conditions**

For the purpose of defining such a hybridizing nucleic acid according to the invention, the stringent hybridization conditions are the followings :

the hybridization step is realized at 65°C in the presence of 6 x SSC buffer, 5 x Denhardt's solution, 0,5% SDS and 100µg/ml of salmon sperm DNA.

The hybridization step is followed by four washing steps :

- two washings during 5 min, preferably at 65°C in a 2 x SSC and 0.1%SDS buffer;
- one washing during 30 min, preferably at 65°C in a 2 x SSC and 0.1% SDS buffer,
- one washing during 10 min, preferably at 65°C in a 0.1 x SSC and 0.1%SDS buffer,

these hybridization conditions being suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The suitable hybridization conditions may for example be adapted according to the teachings disclosed in the book of Hames and Higgins (1985).

## Genomic Sequences Of The *HKLP* Gene

The present invention concerns the genomic sequence of *HKLP* comprising the 2 genomic contigs of SEQ ID Nos 1 and 2. The present invention encompasses *HKLP* gene, or *HKLP* genomic sequenc s consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID Nos 1 and 2, a sequence complementary thereto, as well as fragments and variants thereof. These polynucleotides may be purified, isolated, or recombinant.

The invention also encompasses a purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with a nucleotide sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a complementary sequence thereto or a fragment thereof. The nucleotide differences as regards to the nucleotide sequences of SEQ ID Nos 1 and 2 may be generally randomly distributed throughout the entire nucleic acid. Nevertheless, preferred nucleic acids are those wherein the nucleotide differences as regards to the nucleotide sequences of SEQ ID Nos 1 and 2 are predominantly located outside the coding sequences contained in the exons. These nucleic acids, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the *HKLP* gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the *HKLP* sequences.

Another object of the invention consists of a purified, isolated, or recombinant nucleic acids that hybridizes with a nucleotide sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a complementary sequence thereto or a variant thereof, under the stringent hybridization conditions as defined above.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or 200 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, 10, 20, 30, 40 or 50 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162.

Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884.

Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, or 2, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 nucleotide positions of any one of the following ranges of nucleotide positions of

(a) SEQ ID No 1: 1-1000, 1001-2000, 2001-3000, 3001-4000, 4001-5000, 5001-6000, 6001-7000, 7001-8000, 8001-9000, 9001-10000, 10001-11000, 11001-12000, 12001-13000, 13001-

14000, 14001-15000, 15001-16000, 16001-17000, 17001-18000, 18001-19000, 19001-20000, 20001-21000, 21001-22000, 22001-23000, 23001-24000, 24001-25000, 25001-26000, 26001-27000, 27001-28000, 28001-29000, 29001-30000, 30001-31000, 31001-32000, 32001-33000, 33001-34000, 34001-35000, 35001-36000, 36001-37000, 37001-38000, 38001-39000, 39001-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162; and

(b) SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884.

Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises a G at position 7159 of SEQ ID No 1. Further preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises a C either at position 2551 or 4500 of SEQ ID No 2. It should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section.

The *HKLP* genomic nucleic acid comprises at least 48 exons. The exon positions in SEQ ID Nos 1 and 2 are detailed below in the Table A. The first exon which has been identified in the cDNA of the present invention is not comprised in the genomic sequence described in the present invention. The sequence of the first exon begins at the position 1 of SEQ ID No 3 and ends at the position 292. The genomic sequence of SEQ ID Nos 1 and 2 comprises respectively 44 and 4 exons.

Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the exons of the *HKLP* gene, or a sequence complementary thereto. The invention also deals with purified, isolated, or recombinant nucleic acids comprising a combination of at least two exons of the *HKLP* gene, wherein the polynucleotides are arranged within the nucleic acid, from the 5'-end to the 3'-end of said nucleic acid, in the same order as in SEQ ID Nos 1 and 2.

The position of the introns is detailed in Table A. Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the introns of the *HKLP* gene, or a sequence complementary thereto.

Thus, the present invention deals with a purified or isolated nucleic acid encoding a HKLP protein having the amino acid sequence of SEQ ID No 4 or a peptide fragment or variant thereof. In a specific embodiment, such a purified or isolated nucleic acid comprises a polynucleotide selected

from the group consisting of SEQ ID Nos 1 and 2, or a complementary sequence thereto or a fragment or a variant thereof.

Table A

| Exon | Position in SEQ ID No 1 | | Intron | Position in SEQ ID No 1 | |
|---|---|---|---|---|---|
| | Beginning | End | | Beginning | End |
| 2 | 4844 | 4920 | 2 | 4921 | 7057 |
| 3 | 7058 | 7237 | 3 | 7238 | 10468 |
| 4 | 10469 | 10534 | 4 | 10535 | 15939 |
| 5 | 15940 | 16118 | 5 | 16119 | 16711 |
| 6 | 16712 | 16823 | 6 | 16824 | 20053 |
| 7 | 20054 | 20131 | 7 | 20132 | 20792 |
| 8 | 20793 | 20858 | 8 | 20859 | 21557 |
| 9 | 21558 | 21575 | 9 | 21576 | 23970 |
| 10 | 23971 | 24046 | 10 | 24047 | 24863 |
| 11 | 24864 | 24942 | 11 | 24943 | 26528 |
| 12 | 26529 | 26671 | 12 | 26672 | 27639 |
| 13 | 27640 | 27681 | 13 | 27682 | 30862 |
| 14 | 30863 | 31074 | 14 | 31075 | 39624 |
| 15 | 39625 | 39704 | 15 | 39705 | 40589 |
| 16 | 40590 | 40665 | 16 | 40666 | 43629 |
| 17 | 43630 | 43709 | 17 | 43710 | 44203 |
| 18 | 44204 | 44310 | 18 | 44311 | 45125 |
| 19 | 45126 | 45209 | 19 | 45210 | 45440 |
| 20 | 45441 | 45621 | 20 | 45622 | 45717 |
| 21 | 45718 | 45790 | 21 | 45791 | 68580 |
| 22 | 68581 | 68674 | 22 | 68675 | 70246 |
| 23 | 70247 | 70395 | 23 | 70396 | 72421 |
| 24 | 72422 | 72600 | 24 | 72601 | 73295 |
| 25 | 73296 | 73433 | 25 | 73434 | 74648 |
| 26 | 74649 | 74897 | 26 | 74898 | 83055 |
| 27 | 83056 | 83174 | 27 | 83175 | 85192 |
| 28 | 85193 | 85278 | 28 | 85279 | 85609 |
| 29 | 85610 | 85739 | 29 | 85740 | 85906 |
| 30 | 85907 | 86069 | 30 | 86070 | 88304 |
| 31 | 88305 | 88395 | 31 | 88396 | 90585 |
| 32 | 90586 | 90704 | 32 | 90705 | 91767 |
| 33 | 91768 | 91823 | 33 | 91824 | 94380 |
| 34 | 94381 | 94489 | 34 | 94490 | 96296 |
| 35 | 96297 | 96363 | 35 | 96364 | 97184 |
| 36 | 97185 | 97269 | 36 | 97270 | 101167 |
| 37 | 101168 | 101273 | 37 | 101274 | 109465 |
| 38 | 109466 | 109580 | 38 | 109581 | 110228 |
| 39 | 110229 | 110362 | 39 | 110363 | 111819 |
| 40 | 111820 | 111881 | 40 | 111882 | 113636 |
| 41 | 113637 | 113782 | 41 | 113783 | 113945 |
| 42 | 113946 | 114185 | 42 | 114186 | 117002 |
| 43 | 117003 | 117074 | 43 | 117075 | 119676 |
| 44 | 119677 | 119798 | 44 | 119677 | 121162 |
| Exon | Position in SEQ ID No 2 | | Intron | Position in SEQ ID No 2 | |
| | Beginning | End | | Beginning | End |
| 45 | 1601 | 1750 | 44 | 1 | 1600 |
| 46 | 2139 | 2331 | 45 | 1751 | 2138 |
| 47 | 2540 | 2658 | 46 | 2332 | 2539 |
| 48 | 3830 | 8884 | 47 | 2659 | 3829 |

5

The *HKLP* genomic sequence is covered by two fragments. Indeed, one segment is unknown in the intron 44. The inventors think that this segment, which seems to comprise 20 to 30 nucleotides, forms a superstructure which prevents the sequencing. This superstructure comprises two polyG at each end of the segment.

While this section is entitled "Genomic Sequences of *HKLP*," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of *HKLP* on either side or between two or more such genomic sequences.

## *HKLP* cDNA Sequences

The expression of the *HKLP* gene has been shown to lead to the production of at least one mRNA species, the nucleic acid sequence of which is set forth in SEQ ID No 3.

Another object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID No 3, complementary sequences thereto, as well as allelic variants, and fragments thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant *HKLP* cDNAs consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 3. Particularly preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682. Additional preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises a nucleotide selected in the group consisting of a C at position 5487, and a C at position 6265 of SEQ ID No 3.

The invention also pertains to a purified or isolated nucleic acid having at least 95% of nucleotide identity with the nucleotide sequence of SEQ ID No 3 or a fragment thereof or a complementary sequence thereto, advantageously 99 %, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with the nucleotide sequence of SEQ ID No 3 or a fragment thereof or a complementary sequence thereto.

Another object of the invention consists of a purified, isolated, or recombinant nucleic acids that hybridizes with the nucleotide sequence of SEQ ID No 3 or a complementary sequence thereto or a variant thereof, under the stringent hybridization conditions as defined above.

The cDNA of SEQ ID No 3 includes a 5'-UTR region starting from the nucleotide at position 1 and ending at the nucleotide in position 186 of SEQ ID No 3. The cDNA of SEQ ID No 3 includes a 3'-UTR region starting from the nucleotide at position 5638 and ending at the nucleotide

at position 10682 of SEQ ID No 3. The polyadenylation site starts from the nucleotide at position 10631 and ends at the nucleotide in position 10636 of SEQ ID No 3.

Consequently, the invention concerns a purified, isolated, and recombinant nucleic acids comprising a nucleotide sequence of the 3'UTR of the *HKLP* cDNA, a sequence complementary thereto, or an allelic variant thereof.

While this section is entitled " *HKLP* cDNA Sequences," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of *HKLP* on either side or between two or more such genomic sequences.

## Coding Regions

The *HKLP* open reading frame is contained in the corresponding mRNA of SEQ ID No 3. More precisely, the effective *HKLP* coding sequence (CDS) includes the region between nucleotide position 187 (first nucleotide of the ATG codon) and nucleotide position 5637 (end nucleotide of the TGA codon) of SEQ ID No 3.

The present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4.

The above disclosed polynucleotide that contains the coding sequence of the *HKLP* gene may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the *HKLP* gene of the invention or in contrast the signals may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression and/or amplification.

## Regulatory Sequences Of *HKLP*

As mentioned, the genomic sequence of the *HKLP* gene contains regulatory sequences in the non-coding 3'-flanking region that border the *HKLP* coding region. The 3'-regulatory sequence of the *HKLP* gene is localized between nucleotide position 8885 and nucleotide position 10884 of SEQ ID No 2. Polynucleotides derived from the 3' regulatory region are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID No 2 or a fragment thereof in a test sample.

In order to identify the relevant biologically active polynucleotide fragments or variants of the 3'regulatory region from SEQ ID No 2, the one skill in the art will refer to the book of Sambrook et al.(Sambrook, 1989) which describes the use of a recombinant vector carrying a marker gene (i.e.

beta galactosidase, chloramphenicol acetyl transferase, etc.) the expression of which will be detected when placed under the control of a biologically active polynucleotide fragments or variants of SEQ ID No 2. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a biologically active polynucleotide in the insert.

Polynucleotides carrying the regulatory elements located at the 3' end of the *HKLP* coding region may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest.

Thus, the present invention also concerns a purified or isolated nucleic acid comprising a polynucleotide of the 3' regulatory regions, or a sequence complementary thereto or a biologically active fragment or variant thereof.

Preferred fragments of the 3' regulatory region are at least 50, 100, 150, 200, 300 or 400 bases in length.

By "biologically active" polynucleotide derivatives of SEQ ID No 2 are polynucleotides comprising or alternatively consisting in a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host. It could act either as an enhancer or as a repressor.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are "operably linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

The regulatory polynucleotides of the invention may be prepared from the nucleotide sequence of SEQ ID No 2 by cleavage using suitable restriction enzymes, as described for example in the book of Sambrook et al.(1989). The regulatory polynucleotides may also be prepared by digestion of SEQ ID No 2 by an exonuclease enzyme, such as Bal31 (Wabiko et al., 1986). These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification.

The regulatory polynucleotides according to the invention may be part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

A preferred 3'-regulatory polynucleotide of the invention includes the 3'-untranslated region (3'-UTR) of the *HKLP* cDNA, or a biologically active fragment or variant thereof.

## P lynucleotide C nstructs

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment.

**DNA Construct That Enables Directing Temporal And Spatial *HKLP* Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.**

In order to study the physiological and phenotypic consequences of a lack of synthesis of the HKLP protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the *HKLP* genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to the *HKLP* nucleotide sequence of SEQ ID Nos 1-3, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in an exon of the *HKLP* genomic sequence or within the *HKLP* cDNA of SEQ ID No 3. In a preferred embodiment, the *HKLP* sequence comprises a biallelic marker of the present invention. In a preferred embodiment, the *HKLP* sequence comprises a biallelic marker of the present invention, preferably one of the biallelic markers A1 to A32.

The present invention embodies recombinant vectors comprising any one of the polynucleotides described in the present invention.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn10 for controlling the *HKLP* gene expression, such as described by Gossen et al.(1992, 1995) and Furth et al.(1994). Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to a minimal promoter, said minimal promoter being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a HKLP polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor.

In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

**DNA Constructs Allowing H m l gous Rec mbination: Replacement Vectors**

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the *HKLP* genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the *HKLP* genomic sequence, and is located on the genome downstream the first *HKLP* nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker consists of the thymidine kinase (*tk*) gene (Thomas et al., 1986), the hygromycine beta gene (Te Riele et al., 1990), the *hprt* gene ( Van der Lugt et al., 1991; Reid et al., 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada et al., 1993; Yagi et al.1990). Preferably, the positive selection marker is located within a *HKLP* exon sequence so as to interrupt the sequence encoding a *HKLP* protein. These replacement vectors are described, for example, by Thomas et al.(1986; 1987), Mansour et al.(1988) and Koller et al.(1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a *HKLP* regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

**DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.**

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986). The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique has been first described by Gu et al.(1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al.(1995), or by lipofection of the enzyme into the cells, such as described by Baubonis et al.(1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being

introduced in the recombinant cell host, such as described by Gu et al.(1993) and Sauer et al.(1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu et al.(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the *HKLP* gene by homologous recombination is constructed in such a way that selectable markers are flanked by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the *HKLP* sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are described by Zou et al.(1994).

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the *HKLP* genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the *HKLP* genomic sequence, and is located on the genome downstream of the first *HKLP* nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al.(1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result of the breeding of two transgenic animals, the first transgenic animal bearing the *HKLP*-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al.(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae et al.(1995).

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a *HKLP* genomic sequence or a *HKLP* cDNA sequence, and most preferably an altered copy of a *HKLP* genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination). In a specific embodiment, the DNA constructs described above may be used to introduce a *HKLP* genomic sequence or a *HKLP* cDNA sequence. Optionally, said sequence comprises at least one biallelic marker of the present invention, preferably at least one biallelic marker selected from the group consisting of A1 to A32.

**Nuclear Antisense DNA Constructs**

Other compositions containing a vector of the invention comprising an oligonucleotide fragment of the nucleic sequence SEQ ID No 3, preferably a fragment including the start codon of the *HKLP* gene, as an antisense tool that inhibits the expression of the corresponding *HKLP* gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995) or those described in PCT Application No WO 95/24223.

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the *HKLP* mRNA. In one embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of *HKLP* that contains either the translation initiation codon ATG or a splicing site. Further preferred antisense polynucleotides according to the invention are complementary of the splicing site of the *HKLP* mRNA.

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu et al.(1994). In a preferred embodiment, these *HKLP* antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner et al.(1991).

## Oligonucleotide Probes And Primers

Polynucleotides derived from the *HKLP* gene are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID Nos 1-3, or a fragment, complement, or variant thereof in a test sample.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 a nucleotide of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162. Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884.

Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1or 2, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 nucleotide positions of any one of the following ranges of nucleotide positions of:

(a) SEQ ID No 1: 1-1000, 1001-2000, 2001-3000, 3001-4000, 4001-5000, 5001-6000, 6001-7000, 7001-8000, 8001-9000, 9001-10000, 10001-11000, 11001-12000, 12001-13000, 13001-14000, 14001-15000, 15001-16000, 16001-17000, 17001-18000, 18001-19000, 19001-20000, 20001-21000, 21001-22000, 22001-23000, 23001-24000, 24001-25000, 25001-26000, 26001-27000, 27001-28000, 28001-29000, 29001-30000, 30001-31000, 31001-32000, 32001-33000, 33001-34000, 34001-35000, 35001-36000, 36001-37000, 37001-38000, 38001-39000, 39001-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162; and

(b) SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884.

Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40,

50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises a G at position 7159 of SEQ ID No 1. Further preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises a C either at position 2551 or 4500 of SEQ ID No 2.

Another preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682. Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises a nucleotide selected in the group consisting of a C at position 5487, and a C at position 6265 of SEQ ID No 3.

Thus, the invention also relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a nucleic acid selected from the group consisting of the nucleotide sequences:

a) 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162 of SEQ ID No 1 or a variant thereof or a sequence complementary thereto;

b) 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884 of SEQ ID No 2 or a variant thereof or a sequence complementary thereto; and

c) 391-1619 and 6988-10682 of SEQ ID No 3 or a variant thereof or a sequence complementary thereto.

Additionally, another preferred embodiment of a probe according to the invention consists of a nucleic acid comprising a biallelic marker selected from the group consisting of A1 to A32 or the complements thereto, for which the receptive locations in the sequence listing are provided in Table 2.

The invention also relates to a purified and/or isolated nucleotide sequence comprising a polymorphic base of a *HKLP*-related biallelic marker, preferably of a biallelic marker selected from the group consisting of A1 to A32, and the complements thereof. The sequence has between 8 and 1000 nucleotides in length, and preferably comprises at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 60,

70, 80, 100, 250, 500 or 1000 contiguous nucleotides, to the extent that such lengths are consistent with the specific sequence, of a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-3 and 5-8 or a variant thereof or a complementary sequence thereto. In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of any one of SEQ ID Nos 1-3 and 5-8 and the complement thereof, wherein said span includes a *HKLP*-related biallelic marker in said sequence; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A22 and A25 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; These nucleotide sequences comprise the polymorphic base of either allele 1 or allele 2 of the considered biallelic marker. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of said polynucleotide or at the center of said polynucleotide; optionally, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide; optionally, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide; optionally, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide. Optionally, said polynucleotide may further comprise a label. Optionally, said polynucleotide can be attached to solid support. In a further embodiment, the polynucleotides defined above can be used alone or in any combination. In a preferred embodiment, said probes consists of, or consists essentially of a sequence selected from the following sequences: P1 to P30 and the complementary sequences thereto.

In another embodiment the invention encompasses isolated, purified and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID Nos 1-3 and 5-8 or the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a *HKLP*-related biallelic marker in said sequence, preferably within within 20 nucleotides upstream of a *HKLP*-related biallelic marker in said sequence; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1

to A22 and A25 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said *HKLP*-related biallelic marker in said sequence; and optionally, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D30 and E1 to E30.

In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides comprising, consisting of, or consisting essentially of a sequence selected from the following sequences: B1 to B25 and C1 to C25.

In an additional embodiment, the invention encompasses the use of any polynucleotide for, or polynucleotides for use in determining the identity of the nucleotide at a *HKLP*-related biallelic marker or the complements thereof, as well as polynucleotides for use or use of polynucleotides in amplifying segments of nucleotides comprising a *HKLP*-related biallelic marker or the complements thereof; Optionally, said determining may be performed in hybridization assay, sequencing assays, and enzyme-based mismatch detection assays; Optionally, said amplifying may be performed by a PCR or LCR. optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A22 and A25 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said polynucleotide may be attached to a solid support, array, or addressable array; Optionally, said polynucleotide may be labeled.

The invention concerns the use of the polynucleotides according to the invention for determining the identity of the nucleotide at a *HKLP*-related biallelic marker, preferably in hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay and in amplifying segments of nucleotides comprising a *HKLP*-related biallelic marker. In addition, the polynucleotides of the invention for use or the use of polynucleotides in determining the identity of one or more nucleotides at a *HKLP*-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang et al.(1979), the phosphodiester method of Brown et al.(1979),

the diethylphosphoramidite method of Beaucage et al.(1981) and the solid support method described in EP 0 707 592. The disclosures of all these documents are incorporated herein by reference.

The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

A probe or a primer according to the invention has between 8 and 1000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 nucleotides in length. More particularly, the length of these probes and primers can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art. A preferred probe or primer consists of a nucleic acid comprising a polynucleotide selected from the group of the nucleotide sequences of P1 to P30 and the complementary sequence thereto, B1 to B25, C1 to C25, D1 to D30, E1 to E30, for which the respective locations in the sequence listing are provided in Tables 1, 2, 3 and 4.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances ($^{32}P$, $^{35}S$, $^{3}H$, $^{125}I$), fluorescent dyes (5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling

of nucleic acid fragments are described in the French patent No. FR-7810975 or by Urdea et al (1988) or Sanchez-Pescador et al (1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the *HKLP* gene or mRNA using other techniques.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to

the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Consequently, the invention also deals with a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-3, a fragment or a variant thereof and a complementary sequence thereto in a sample, said method comprising the following steps of:

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected form the group consisting of the nucleotide sequences of SEQ ID Nos 1-3, a fragment or a variant thereof and a complementary sequence thereto and the sample to be assayed; and

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-3, a fragment or a variant thereof and a complementary sequence thereto in a sample, said kit comprising:

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected form the group consisting of the nucleotide sequences of SEQ ID Nos 1-3, a fragment or a variant thereof and a complementary sequence thereto; and

b) optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of this detection method and kit, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule. In a second preferred embodiment of said method and kit, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate. In a third preferred embodiment, the nucleic acid probe or the plurality of nucleic acid probes comprise either a sequence which is selected from the group consisting of the nucleotide sequences of P1 to P30 and the complementary sequence thereto, B1 to B25, C1 to C25, D1 to D30, E1 to E30 or a biallelic marker selected from the group consisting of A1 to A32 and the complements thereto.

**Olig nucleotide Arrays**

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in the *HKLP* gene and may also be used for detecting mutations in the coding or in the non-coding sequences of the *HKLP* gene.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in the *HKLP* gene and in its regulatory region. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the *HKLP* gene that have been identified according, for example to the technique used by Huang et al.(1996) or Samson et al.(1996).

Another technique that is used to detect mutations in the *HKLP* gene is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the *HKLP* genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the *HKLP* gene. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. This technique was described by Chee et al. in 1996, which is herein incorporated by reference.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

A further object of the invention consists of an array of nucleic acid sequences comprising either at least one of the sequences selected from the group consisting of P1 to P30, B1 to B25, C1 to C25, D1 to D30, E1 to E30, the sequences complementary thereto, a fragment thereof of at least 8 consecutive nucleotides thereof, and at least one sequence comprising a biallelic marker selected from the group consisting of A1 to A32 and the complements thereto.

The invention also pertains to an array of nucleic acid sequences comprising either at least two of the sequences selected from the group consisting of P1 to P30, B1 to B25, C1 to C25, D1 to D30, E1 to E30, the sequences complementary thereto, a fragment thereof of at least 8 consecutive nucleotides thereof, and at least two sequences comprising a biallelic marker selected from the group consisting of A1 to A32 and the complements thereof.

## Amplification of the *HKLP* gene.

### 1. DNA extraction

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples which can be tested by the methods of the present invention described herein and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed

tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the context of the present invention is from peripheral venous blood of each donor.

The techniques of DNA extraction are well-known to the skilled technician. Such techniques are described notably by Mackey et al. (1998).

**2. DNA amplification**

DNA amplification techniques are well-known to those skilled in the art. Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the disclosures of which are incorporated herein by reference, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli JC, et al. (1990) and in Compton J. (1991), Q-beta amplification as described in European Patent Application no 4544610, strand displacement amplification as described in Walker et al. (1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461, the disclosure of which is incorporated herein by reference.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR

(RT-AGLCR) as described by Marshall et al. (1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188. Each of these publications is incorporated by reference.

One of the aspects of the present invention is a method for the amplification of the human *HKLP* gene, particularly of the genomic sequences of SEQ ID No 1 and 2 or of the cDNA sequence of SEQ ID No 3, or a fragment or a variant thereof in a test sample, preferably using the PCR technology. The method comprises the steps of contacting a test sample suspected of containing the target *HKLP* encoding sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers, and eventually in some instances a detection probe that can hybridize with an internal region of amplicon sequences to confirm that the desired amplification reaction has taken place.

Thus, the present invention also relates to a method for the amplification of a human *HKLP* gene sequence, particularly of a portion of the genomic sequences of SEQ ID Nos 1 and 2 or of the cDNA sequence of SEQ ID No 3, or a variant thereof in a test sample, said method comprising the steps of:

a) contacting a test sample suspected of containing the targeted *HKLP* gene sequence comprised in a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-3, or fragments or variants thereof with amplification reaction reagents comprising a pair of amplification primers as described above and located on either side of the polynucleotide region to be amplified; and

b) optionally, detecting the amplification products.

The invention also concerns a kit for the amplification of a human *HKLP* gene sequence, particularly of a portion of the genomic sequences of SEQ ID No 1 and 2 or of the cDNA sequence of SEQ ID No 3, or a variant thereof in a test sample, wherein said kit comprises:

a) a pair of oligonucleotide primers located on either side of the *HKLP* region to be amplified; and

b) Optionally, the reagents necessary for performing the amplification reaction.

In a first preferred embodiment of the above amplification method or kit, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region.

The primers are more particularly characterized in that they have sufficient complementarity with any sequence of a strand of the genomic sequence close to the region to be amplified, for example with a non-coding sequence adjacent to exons to amplify.

In a second preferred embodiment, the nucleic acid primers comprise a sequence which is selected from the group consisting of the nucleotide sequences of B1 to B25, C1 to C25, D1 to D30, and E1 to E30.

**HKLP Proteins and Polypeptide Fragments:**

The term "HKLP polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies HKLP proteins from humans, including isolated or purified HKLP proteins consisting, consisting essentially, or comprising the sequence of SEQ ID No 4. The HKLP protein has 1816 amino acids in length. The 700 first amino acids of the HKLP protein present 97 % of homology with the murine KIF1B protein (Nangaku et al., 1994). The HKLP protein presents 60-70% of homology with the murine KIF1A protein, and more particularly the 390 first amino acids of the HKLP protein have 85 % of homology therewith.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the HKLP protein sequence.

The invention also encompasses a purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, 95, 98 or 99% nucleotide identity with a nucleotide sequence of SEQ ID No 4 or a complementary sequence thereto or a fragment thereof.

HKLP proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes. The HKLP polypeptides of the invention can be made using routine expression methods known in the art. The polynucleotide encoding the desired

polypeptide is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides, and a summary of some of the more common systems. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

In addition, shorter protein fragments is produced by chemical synthesis. Alternatively the proteins of the invention is extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Any *HKLP* cDNA, including SEQ ID No 3, is used to express HKLP proteins and polypeptides. The nucleic acid encoding the HKLP protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The *HKLP* insert in the expression vector may comprise the full coding sequence for the HKLP protein or a portion thereof. For example, the *HKLP* derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the HKLP protein of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

In one embodiment, the entire coding sequence of the *HKLP* cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the HKLP protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the *HKLP* cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene.

The nucleic acid encoding the HKLP protein or a portion thereof is obtained by PCR from a bacterial vector containing the *HKLP* cDNA of SEQ ID No 3 using oligonucleotide primers complementary to the *HKLP* cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the HKLP protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

Alternatively, the nucleic acids encoding the HKLP protein or a portion thereof is cloned into pED6dpc2 (Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs is transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded.

The above procedures may also be used to express a mutant HKLP protein responsible for a detectable phenotype or a portion thereof.

The expressed proteins are purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed HKLP protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the HKLP protein or portion thereof is attached to the chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the HKLP protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the HKLP protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the HKLP protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the HKLP protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed HKLP protein or a portion thereof are described below.

If antibody production is not possible, the nucleic acids encoding the HKLP protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the HKLP protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the β-globin gene or the nickel binding polypeptide and the HKLP protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

Thus, the present invention also concerns a method for producing one of the polypeptides described herein, and especially a polypeptide of SEQ ID No 4 or a fragment or a variant thereof, wherein said method comprises the steps of :

a) culturing, in an appropriate culture medium, a cell host previously transformed or transfected with the recombinant vector comprising a nucleic acid encoding a HKLP polypeptide, or a fragment or a variant thereof;

b) harvesting the culture medium thus conditioned or lyse the cell host, for example by sonication or by an osmotic shock;

c) separating or purifying, from the said culture medium, or from the pellet of the resultant host cell lysate the thus produced polypeptide of interest.

d) Optionally characterizing the produced polypeptide of interest.

In a specific embodiment of the above method, step a) is preceded by a step wherein the nucleic acid coding for a HKLP polypeptide, or a fragment or a variant thereof, is inserted in an appropriate vector, optionally after an appropriate cleavage of this amplified nucleic acid with one or several restriction endonucleases. The nucleic acid coding for a HKLP polypeptide or a fragment or a variant thereof may be the resulting product of an amplification reaction using a pair of primers according to the invention (by SDA, TAS, 3SR NASBA, TMA etc.).

**Antibodies That Bind HKLP Polypeptides of the Invention**

Any HKLP polypeptide or whole protein may be used to generate antibodies capable of specifically binding to expressed HKLP protein or fragments thereof as described. The antibody compositions of the invention are capable of specifically binding or specifically bind to the HKLP protein. For an antibody composition to specifically bind to the HKLP protein it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for full length HKLP protein than for any full length protein in an ELISA, RIA, or other antibody-based binding assay.

In a preferred embodiment of the invention antibody compositions are capable of selectively binding, or selectively bind to an epitope-containing fragment of a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said epitope comprises at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4, wherein said antibody composition is optionally either polyclonal or monoclonal.

The present invention also contemplates the use of polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of a HKLP polypeptide in the manufacture of antibodies, wherein said contiguous span comprises at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4. In a preferred embodiment such polypeptides are useful in the manufacture of antibodies to detect the presence and absence of the HKLP protein.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of HKLP than the one to which antibody binding is desired, and animals which do not express HKLP (i.e. a HKLP knock out animal as described in herein) are particularly useful for preparing antibodies. HKLP knock out animals will recognize all or most of the exposed regions of HKLP as foreign antigens, and therefore produce antibodies with a wider array of HKLP epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to the HKLP protein. In addition, the humoral immune system of animals which produce a species of HKLP that resembles the antigenic sequence will preferentially recognize the differences between the animal's native HKLP species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to the HKLP protein.

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

The antibodies of the invention may be labeled, either by a radioactive, a fluorescent or an enzymatic label.

Consequently, the invention is also directed to a method for detecting specifically the presence of a human HKLP polypeptide according to the invention in a biological sample, said method comprising the following steps :

a) bringing into contact the biological sample with a polyclonal or monoclonal antibody directed against the HKLP polypeptide of the amino acid sequence of SEQ ID No 4, or to a peptide fragment or variant thereof;

b) detecting the antigen-antibody complex formed.

The invention also concerns a diagnostic kit for detecting *in vitro* the presence of a human HKLP polypeptide according to the present invention in a biological sample, wherein said kit comprises :

a) a polyclonal or monoclonal antibody directed against the HKLP polypeptide of the amino acid sequence of SEQ ID No 4, or to a peptide fragment or variant thereof, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

## *HKLP* -related Biallelic Markers

### Advantages Of The Biallelic Markers Of The Present Invention

The *HKLP*-related biallelic markers of the present invention offer a number of important advantages over other genetic markers such as RFLP (Restriction fragment length polymorphism) and VNTR (Variable Number of Tandem Repeats) markers.

The first generation of markers, were RFLPs, which are variations that modify the length of a restriction fragment. But methods used to identify and to type RFLPs are relatively wasteful of materials, effort, and time. The second generation of genetic markers were VNTRs, which can be categorized as either minisatellites or microsatellites. Minisatellites are tandemly repeated DNA sequences present in units of 5-50 repeats which are distributed along regions of the human chromosomes ranging from 0.1 to 20 kilobases in length. Since they present many possible alleles, their informative content is very high. Minisatellites are scored by performing Southern blots to identify the number of tandem repeats present in a nucleic acid sample from the individual being tested. However, there are only $10^4$ potential VNTRs that can be typed by Southern blotting. Moreover, both RFLP and VNTR markers are costly and time-consuming to develop and assay in large numbers.

Single nucleotide polymorphism or biallelic markers can be used in the same manner as RFLPs and VNTRs but offer several advantages. SNP are densely spaced in the human genome and represent the most frequent type of variation. An estimated number of more than $10^7$ sites are

scattered along the $3x10^9$ base pairs of the human genome. Therefore, SNP occur at a greater frequency and with greater uniformity than RFLP or VNTR markers which means that there is a greater probability that such a marker will be found in close proximity to a genetic locus of interest. SNP are less variable than VNTR markers but are mutationally more stable.

Also, the different forms of a characterized single nucleotide polymorphism, such as the biallelic markers of the present invention, are often easier to distinguish and can therefore be typed easily on a routine basis. Biallelic markers have single nucleotide based alleles and they have only two common alleles, which allows highly parallel detection and automated scoring. The biallelic markers of the present invention offer the possibility of rapid, high throughput genotyping of a large number of individuals.

Biallelic markers are densely spaced in the genome, sufficiently informative and can be assayed in large numbers. The combined effects of these advantages make biallelic markers extremely valuable in genetic studies. Biallelic markers can be used in linkage studies in families, in allele sharing methods, in linkage disequilibrium studies in populations, in association studies of case-control populations or of trait positive and trait negative populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in complex traits. Association studies examine the frequency of marker alleles in unrelated case- and control-populations and are generally employed in the detection of polygenic or sporadic traits. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies). Biallelic markers in different genes can be screened in parallel for direct association with disease or response to a treatment. This multiple gene approach is a powerful tool for a variety of human genetic studies as it provides the necessary statistical power to examine the synergistic effect of multiple genetic factors on a particular phenotype, drug response, sporadic trait, or disease state with a complex genetic etiology.

**Candidate Gene Of The Present Invention**

Different approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. Genome-wide association studies rely on the screening of genetic markers evenly spaced and covering the entire genome. The candidate gene approach is based on the study of genetic markers specifically located in genes potentially involved in a biological pathway related to the trait of interest. In the present invention, *HKLP* is the candidate gene. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. However, it should be noted that all of the biallelic markers disclosed in the instant application can be employed

as part of genome-wide association studies or as part of candidate region association studies and such uses are specifically contemplated in the present invention and claims.

***HKLP*-Related Biallelic Markers And P lynucleotides Related Thereto**

The invention also concerns *HKLP*-related biallelic markers. As used herein the term "*HKLP*-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with the *HKLP* gene. The term *HKLP*-related biallelic marker includes the biallelic markers designated A1 to A32.

A portion of the biallelic markers of the present invention are disclosed in Table 2. Their location on the *HKLP* gene is indicated in Table 2 and also as a single base polymorphism in the features of in the related SEQ ID Nos 1-3 and 5-8. The pairs of primers allowing the amplification of a nucleic acid containing the polymorphic base of one *HKLP* biallelic marker are listed in Table 1 of Example 2.

27 *HKLP*-related biallelic markers, A1 to A27, are located in the genomic sequence of *HKLP*. Four of them are located in exonic sequence, namely A1, A23, A24 and A25. The other *HKLP*-related biallelic markers are located in intronic region of *HKLP*. Additionally, 5 biallelic markers are located in intergenic region and are in linkage desiquilibirum with the *HKLP* gene.

The primers for amplification or sequencing reaction of a polynucleotide comprising a biallelic marker of the invention may be designed from the disclosed sequences for any method known in the art. A preferred set of primers are fashioned such that the 3' end of the contiguous span of identity with a sequence selected from the group consisting of SEQ ID Nos 1-3 and 5-8 or a sequence complementary thereto or a variant thereof is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. Allele specific primers may be designed such that a polymorphic base of a biallelic marker is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker. The 3' end of the primer of the invention may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a *HKLP*-related biallelic marker in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers. Thus, another set of preferred amplification primers comprise an isolated polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides in a sequence selected from the group consisting of SEQ ID Nos 1-3 and 5-8 or a sequence complementary thereto or a variant thereof, wherein the 3' end of said contiguous span is located at the 3'end of said polynucleotide, and wherein the 3'end of said polynucleotide is located upstream of a *HKLP*-related biallelic marker in said sequence. Preferably, those

amplification primers comprise a sequence selected from the group consisting of the sequences B1 to B25 and C1 to C25. Primers with their 3' ends located 1 nucleotide upstream of a *HKLP*-related biallelic marker have a special utility in microsequencing assays. Preferred microsequencing primers are described in Table 4. Optionally, microsequencing primers are selected from the group consisting of the nucleotide sequences D1 to D30 and E1 to E30.

The probes of the present invention may be designed from the disclosed sequences for any method known in the art, particularly methods which allow for testing if a marker disclosed herein is present. A preferred set of probes may be designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic marker, but not the other allele under any particular set of assay conditions. Preferred hybridization probes comprise the polymorphic base of either allele 1 or allele 2 of the specific biallelic marker. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe.

It should be noted that the polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases which are enumerated in Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers may be lengthened or shortened to any extent compatible with their intended use and the present invention specifically contemplates such sequences. The flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in human subjects. The addition of any nucleotide sequence which is compatible with the nucleotides intended use is specifically contemplated.

Primers and probes may be labeled or immobilized on a solid support as described in "Oligonucleotide probes and primers".

The polynucleotides of the invention which are attached to a solid support encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said polynucleotides may be specified as attached individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. Optionally, polynucleotides other than those of the invention may attached to the same solid support as polynucleotides of the invention. Optionally, when multiple polynucleotides are attached to a solid support they may be attached at random locations, or in an ordered array. Optionally, said ordered array may be addressable.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at a *HKLP*-related biallelic marker. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not

limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or an enzyme-based mismatch detection method.

## Methods For *De Novo* Identification Of Biallelic Markers

Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms such as differential hybridization with oligonucleotide probes, detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

In a first embodiment, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained thereby usually demonstrates a sufficient frequency of its less common allele to be useful in conducting association studies.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. Such a biallelic marker will, however, be sufficiently informative to conduct association studies and it will further be appreciated that including less informative biallelic markers in the genetic analysis studies of the present invention, may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

### Genomic DNA Samples

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. Details of a preferred embodiment are provided in Example 1. The person skilled in the art can choose to amplify pooled or unpooled DNA samples.

**DNA Amplification**

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art. Various methods to amplify DNA fragments carrying biallelic markers are further described hereinbefore in "Amplification of the *HKLP* gene". The PCR technology is the preferred amplification technique used to identify new biallelic markers. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 2.

In a first embodiment of the present invention, biallelic markers are identified using genomic sequence information generated by the inventors. Sequenced genomic DNA fragments are used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

Preferred primers, useful for the amplification of genomic sequences encoding the candidate genes, focus on promoters, exons and splice sites of the genes. A biallelic marker presents a higher probability to be an eventual causal mutation if it is located in these functional regions of the gene. Preferred amplification primers of the invention include the nucleotide sequences B1 to B25 and C1 to C25, detailed further in Example 2, Table 1.

**Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms**

The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either

the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are for example disclosed in Sambrook et al.(1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al.(1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis. The polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. Because each dideoxy terminator is labeled with a different fluorescent molecule, the two peaks corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to be registered as a polymorphic sequence, the polymorphism has to be detected on both strands.

The above procedure permits those amplification products, which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90% of the biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele. Preferably at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than 0.7 for the major allele, thus a heterozygosity rate higher than 0.18, preferably higher than 0.32, more preferably higher than 0.42.

In another embodiment, biallelic markers are detected by sequencing individual DNA samples, the frequency of the minor allele of such a biallelic marker may be less than 0.1.

**Validation Of The Biallelic Markers Of The Present Invention**

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals,

so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

**Evaluation Of The Frequency Of The Biallelic Markers Of The Present Invention**

The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The higher the frequency of the less common allele the greater the usefulness of the biallelic marker is association and interaction studies. The determination of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. This determination of frequency by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with high quality biallelic markers.

The invention also relates to methods of estimating the frequency of an allele in a population comprising: a) genotyping individuals from said population for said biallelic marker according to the method of the present invention; b) determining the proportional representation of said biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A22 and A25 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24; optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by determining the identity of the nucleotides for both

copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said nucleotide at said *HKLP*-related biallelic marker for the population; optionally, determining the proportional representation may be accomplished by performing a genotyping method of the invention on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and calculating the proportional amount of said nucleotide compared with the total.

## Methods For Genotyping An Individual For Biallelic Markers

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at a *HKLP* biallelic marker site by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed on nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

In one embodiment the invention encompasses methods of genotyping comprising determining the identity of a nucleotide at a *HKLP*-related biallelic marker or the complement thereof in a biological sample; Optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A17, and A20 to A22, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said biological sample is derived from a single subject; optionally, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; optionally, wherein said biological sample is derived from multiple subjects; optionally, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said

determining step; optionally, wherein said amplifying is performed by PCR; optionally, wherein said determining is performed by a hybridization assay, a sequencing assay, a microsequencing assay, or an enzyme-based mismatch detection assay.

**Source of DNA for genotyping**

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

**Amplification Of DNA Fragments Comprising Biallelic Markers**

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention. Amplification of DNA may be achieved by any method known in the art. Amplification techniques are described above in the section entitled, "Amplification of the *HKLP* gene".

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described below.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Example 2. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present inv ntion, amplified segments carrying biallelic markers can range in size

from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described in "Oligonucleotide probes and primers".

**Methods of Genotyping DNA samples for Biallelic Markers**

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al.(1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield et al.(1991), White et al.(1992), Grompe et al.(1989 and 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing Assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described in "Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms".

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

2) Microsequencing Assays

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 4.

Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and Chen et al.(1997). In this method, amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogeneous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator regent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if

more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner, oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, 1994) or linked to fluorescein (Livak and Hainer, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al.(1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al.(1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include the nucleotide sequences D1 to D30 and E1 to E30. It will be appreciated that the microsequencing primers listed in Example 4 are merely exemplary and that, any primer having a 3' end immediately adjacent to the polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Example 4, or fragments comprising at least 8, 12, 15, 20, 25, 30, 40, or 50 consecutive nucleotides thereof and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at a biallelic marker site.

3) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on

correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in "Amplification Of DNA Fragments Comprising Biallelic Markers".

**Allele Specific Amplification Primers**

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This is accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Determining the precise location of the mismatch and the corresponding assay conditions are well with the ordinary skill in the art.

**Ligation/Amplification Based Methods**.

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and may be advantageously combined with PCR as described by Nickerson et al.(1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other amplification methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in "Amplification of the *HKLP* gene". LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO

90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

4) Hybridization Assay Methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. Although such hybridizations can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization

duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., 1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., 1995). In an alternative homogeneous hybridization based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., 1998).

The polynucleotides provided herein can be used to produce probes which can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes, the biallelic marker is at the center of said polynucleotide. Preferred probes comprise a nucleotide sequence selected from the group consisting of amplicons listed in Table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. In preferred embodiments the polymorphic base is within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more preferably at the center of said polynucleotide.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in "Oligonucleotide Probes and Primers". The probes can be non-extendable as described in "Oligonucleotide Probes and Primers".

By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample. High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

5) Hybridization To Addressable Arrays Of Oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (e.g., the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., 1996; Shoemaker et al., 1996; Kozal et al., 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP 785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular, the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in

the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of amplicons listed in table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. In preferred embodiments the polymorphic base is within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more preferably at the center of said polynucleotide. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in "oligonucleotide probes and primers".

6) Integrated Systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

## Methods Of Genetic Analysis Using The Biallelic Markers Of The Present Invention

Different methods are available for the genetic analysis of complex traits (see Lander and Schork, 1994). The search for disease-susceptibility genes is conducted using two main methods: the linkage approach in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, and the association approach in which evidence is sought for a statistically significant association between an allele and a trait or a trait causing allele (Khoury et al., 1993). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant correlation between a genotype and a phenotype. The biallelic markers may be used in parametric and non-parametric linkage analysis methods.

Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach which does not require the use of affected families and which permits the identification of genes associated with complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention may be conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention corresponding to the candidate gene may be used. Further, any set of genetic markers including a biallelic marker of the present invention may be used. A set of biallelic polymorphisms that could be used as genetic markers in combination with the biallelic markers of the present invention has been described in WO 98/20165. As mentioned above, it should be noted that the biallelic markers of the present invention may be included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

**Linkage Analysis**

Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Thus, the aim of linkage analysis is to detect marker loci that show cosegregation with a trait of interest in pedigrees.

Parametric Methods

When data are available from successive generations there is the opportunity to study the degree of linkage between pairs of loci. Estimates of the recombination fraction enable loci to be ordered and placed onto a genetic map. With loci that are genetic markers, a genetic map can be established, and then the strength of linkage between markers and traits can be calculated and used to indicate the relative positions of markers and genes affecting those traits (Weir, 1996). The classical method for linkage analysis is the logarithm of odds (lod) score method (see Morton, 1955; Ott, 1991). Calculation of lod scores requires specification of the mode of inheritance for the disease (parametric method). Generally, the length of the candidate region identified using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (i.e., the ratio between the number of trait positive carriers of allele *a* and the total number of *a* carriers in the population). However, parametric linkage analysis suffers from a variety of drawbacks. First, it is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to

refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, parametric linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due t the combined action of multiple genes and/or environmental factors. It is very difficult to model these factors adequately in a lod score analysis. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (1996).

Non-Parametric Methods

The advantage of the so-called non-parametric methods for linkage analysis is that they do not require specification of the mode of inheritance for the disease, they tend to be more useful for the analysis of complex traits. In non-parametric methods, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Affected relatives should show excess "allele sharing" even in the presence of incomplete penetrance and polygenic inheritance. In non-parametric linkage analysis the degree of agreement at a marker locus in two individuals can be measured either by the number of alleles identical by state (IBS) or by the number of alleles identical by descent (IBD). Affected sib pair analysis is a well-known special case and is the simplest form of these methods.

The biallelic markers of the present invention may be used in both parametric and non-parametric linkage analysis. Preferably biallelic markers may be used in non-parametric methods which allow the mapping of genes involved in complex traits. The biallelic markers of the present invention may be used in both IBD- and IBS- methods to map genes affecting a complex trait. In such studies, taking advantage of the high density of biallelic markers, several adjacent biallelic marker loci may be pooled to achieve the efficiency attained by multi-allelic markers (Zhao et al., 1998).

**Population Association Studies**

The present invention comprises methods for identifying if the *HKLP* gene is associated with a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. In a preferred embodiment, the biallelic markers of the present invention are used to perform candidate gene association studies. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait

when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention may be incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers has been described in US Provisional Patent application serial number 60/082,614. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal segment for example).

As mentioned above, association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families. Association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis methods. Moreover, once a chromosome segment of interest has been identified, the presence of a candidate gene such as a candidate gene of the present invention, in the region of interest can provide a shortcut to the identification of the trait causing allele. Biallelic markers of the present invention can be used to demonstrate that a candidate gene is associated with a trait. Such uses are specifically contemplated in the present invention.

**Determining The Frequency Of A Biallelic Marker Allele Or Of A Biallelic Marker Haplotype In A Population**

Association studies explore the relationships among frequencies for sets of alleles between loci.

Determining The Frequency Of An Allele In A Population

Allelic frequencies of the biallelic markers in a populations can be determined using one of the methods described above under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. Genotyping pooled samples or individual samples can determine the frequency of a biallelic marker allele in a population. One way to reduce the number of genotypings required is to use pooled samples. A major obstacle in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple gene counting is applied to determine the frequency of an allele of a biallelic marker or of a genotype in a given population.

Determining Th Frequency Of A Haplotype In A Population

The gametic phase of haplotypes is unknown when diploid individuals are heterozygous at more than one locus. Using genealogical information in families gametic phase can sometimes be inferred (Perlin et al., 1994). When no genealogical information is available different strategies may be used. One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR amplification (see Newton et al, 1989; Wu et al., 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano et al., 1990). Further, a sample may be haplotyped for sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S. S., 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalization at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm to infer the phase of PCR-amplified DNA genotypes introduced by Clark, A.G.(1990) may be used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognized haplotypes. For each positive identification, the complementary haplotype is added to the list of recognized haplotypes, until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Alternatively, one can use methods estimating haplotype frequencies in a population without assigning haplotypes to each individual. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster et al., 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used (see Excoffier L. and Slatkin M., 1995). The EM algorithm is a generalized iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical Methods." Any other method known in the art to determine or to estimate the frequency of a haplotype in a population may be used.

The invention also encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping at least one *HKLP*-related biallelic marker according to a method of the invention for each individual in said population; b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copi s of said second biallelic marker present in the genome of

each individual in said population; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A17, and A20 to A22, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said haplotype determination method is performed by asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark algorithm, or an expectation-maximization algorithm.

**Linkage Disequilibrium Analysis**

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for mapping genes involved in disease traits (see Ajioka R.S. et al., 1997). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in greater numbers than other types of genetic markers (such as RFLP or VNTR markers), are particularly useful in genetic analysis based on linkage disequilibrium.

When a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombination events occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those that are further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations. Linkage disequilibrium is usually seen as an association between one specific allele at one locus and another specific allele at a second locus.

The pattern or curve of disequilibrium between disease and marker loci is expected to exhibit a maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between

markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The high density of biallelic markers combined with linkage disequilibrium analysis provides powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods".

**Population-Based Case-Control Studies Of Trait-Marker Associations**

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random and the deviation from random is called linkage disequilibrium. Association studies focus on population frequencies and rely on the phenomenon of linkage disequilibrium. If a specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (trait positive) population, when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele will also be increased in trait positive individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular region. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. As any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

Case-Control Populations (Inclusion Criteria)

Population-based association studies do not concern familial inheritance but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (unaffected, trait negative or random) individuals. Preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is ethnically matched to the case population. Moreover, the control group is preferably matched to the case-population for the main known confusion factor for the trait under study (for example age-matched for an age-dependent trait). Ideally, individuals in the two samples are paired in such a way that they are expected to differ only in their disease status. The terms "trait positive population", "case population" and "affected population" are used interchangeably herein.

An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, 1994). A major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Any genetic trait may be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait positive and trait negative phenotypic groups. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. The selection procedure for continuous or quantitative traits (such as blood pressure for example) involves selecting individuals at opposite ends of the phenotype distribution of the trait under study, so as to include in these trait positive and trait negative populations individuals with non-overlapping phenotypes. Preferably, case-control populations consist of phenotypically homogeneous populations. Trait positive and trait negative populations consist of phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably between 1 and 30%, most preferably between 1 and 20% of the total population under study, and preferably selected among individuals exhibiting non-overlapping phenotypes. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

In preferred embodiments, a first group of between 50 and 300 trait positive individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of control individuals are included in such studies.

Association Analysis

The invention also comprises methods of detecting an association between a genotype and a phenotype, comprising the steps of: a) determining the frequency of at least one *HKLP*-related biallelic marker in a trait positive population according to a genotyping method of the invention; b) determining the frequency of said *HKLP*-related biallelic marker in a control population according to a genotyping method of the invention; and c) determining whether a statistically significant association exists between said genotype and said phenotype. In addition, the methods of detecting an association between a genotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A17, and A20 to A22, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the

biallelic markers in linkage disequilibrium therewith; Optionally, said control population may be a trait negative population, or a random population; Optionally, each of said genotyping steps a) and b) may be performed on a pooled biological sample derived from each of said populations; Optionally, each of said genotyping of steps a) and b) is performed separately on biological samples derived from each individual in said population or a subsample thereof.

The general strategy to perform association studies using biallelic markers derived from a region carrying a candidate gene is to scan two groups of individuals (case-control populations) in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (i.e. the associated allele is the trait causing allele), or more likely the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele with respect to the candidate gene function usually give further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker, and performing further association studies with the polymorphisms that are revealed in an iterative manner.

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers from the candidate gene are determined in the trait positive and control populations. In a second phase of the analysis, the position of the genetic loci responsible for the given trait is further refined using a higher density of markers from the relevant region. However, if the candidate gene under study is relatively small in length, as is the case for *HKLP*, a single phase may be sufficient to establish significant associations.

Haplotype Analysis

As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analyzed. Complementing single point (allelic) association studies with multi-point association studies also called haplotype studies increases the statistical power of association studies. Thus, a haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical power of an analysis involving individual markers.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined.

The haplotype frequency is then compared for distinct populations of trait positive and control individuals. The number of trait positive individuals, which should be, subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of individuals ranging between 50 and 150. The same considerations apply to the number of unaffected individuals (or random control) used in the study. The results of this first analysis provide haplotype frequencies in case-control populations, for each evaluated haplotype frequency a p-value and an odd ratio are calculated. If a statistically significant association is found the relative risk for an individual carrying the given haplotype of being affected with the trait under study can be approximated.

An additional embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a trait positive population, according to a method of the invention for estimating the frequency of a haplotype; b) estimating the frequency of said haplotype in a control population, according to a method of the invention for estimating the frequency of a haplotype; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A17, and A20 to A22, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said control population is a trait negative population, or a random population. Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c).

Interaction Analysis

The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set of biallelic markers with appropriate level of statistical significance can be considered as a haplotype analysis. Interaction analysis consists in stratifying the case-control populations with respect to a given haplotype for the first loci and performing a haplotype analysis with the second loci with each subpopulation.

Statistical methods used in association studies are further described below.

**Testing For Linkage In The Presence Of Association**

The biallelic markers of the present invention may further be used in TDT (transmission/disequilibrium test). TDT tests for both linkage and association and is not affected by population stratification. TDT requires data for affected individuals and their parents or data from unaffected sibs instead of from parents (see Spielmann S. et al., 1993; Schaid D.J. et al., 1996, Spielmann S. and Ewens W.J., 1998). Such combined tests generally reduce the false – positive errors produced by separate analyses.

**Statistical methods**

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation may be used.

**1) Methods In Linkage Analysis**

Statistical methods and computer programs useful for linkage analysis are well-known to those skilled in the art (see Terwilliger J.D. and Ott J., 1994; Ott J., 1991).

**2) Methods To Estimate Haplotype Frequencies In A Population**

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., 1997; Weir, B.S., 1996) Preferably, maximum-likelihood haplotype frequencies are computed using an Expectation- Maximization (EM) algorithm (see Dempster et al., 1977; Excoffier L. and Slatkin M., 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M. E. et al., 1994) or the Arlequin program (Schneider et al., 1997). The EM algorithm is a generalized iterative maximum likelihood approach to estimation and is briefly described below.

Please note that in the present section, "Methods To Estimate Haplotype Frequencies In A Population, " of this text, phenotypes will refer to multi-locus genotypes with unknown phase. Genotypes will refer to known-phase multi-locus genotypes.

A sample of N unrelated individuals is typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can categorized in F different phenotypes. Suppose that we have H underlying possible haplotypes (in case of K biallelic markers, $H=2^K$).

For phenotype j, suppose that $c_j$ genotypes are possible. We thus have the following equation

$$P_j = \sum_{i=1}^{c_j} pr(genotype_i) = \sum_{i=1}^{c_j} pr(h_k, h_l) \qquad \text{Equation 1}$$

where $Pj$ is the probability of the phenotype $j$, $h_k$ and $h_l$ are the two haplotypes constituent the genotype $i$. Under the Hardy-Weinberg equilibrium, $pr(h_k,h_l)$ becomes:

$$pr(h_k, h_l) = pr(h_k)^2 \text{ if } h_k = h_l,\ pr(h_k, h_l) = 2pr(h_k).pr(h_l) \text{ if } h_k \neq h_l. \quad \text{Equation 2}$$

The successive steps of the E-M algorithm can be described as follows:

Starting with initial values of the of haplotypes frequencies, noted $p_1^{(0)}, p_2^{(0)}, \ldots . p_H^{(0)}$, these initial values serve to estimate the genotype frequencies (Expectation step) and then estimate another set of haplotype frequencies (Maximization step), noted $p_1^{(1)}, p_2^{(1)}, \ldots . p_H^{(1)}$, these two steps are iterated until changes in the sets of haplotypes frequency are very small.

A stop criterion can be that the maximum difference between haplotype frequencies between two iterations is less than $10^{-7}$. These values can be adjusted according to the desired precision of estimations.

At a given iteration s, the Expectation step consists in calculating the genotypes frequencies by the following equation:

$$pr(genotype_i)^{(s)} = pr(phenotype_j).pr(genotype_i | phenotype_j)^{(s)}$$

$$= \frac{n_j}{N} \cdot \frac{pr(h_k, h_l)^{(s)}}{P_j^{(s)}} \qquad \text{Equation 3}$$

where genotype $i$ occurs in phenotype $j$, and where $h_k$ and $h_l$ constitute genotype $i$. Each probability is derived according to eq. 1, and eq. 2 described above.

Then the Maximization step simply estimates another set of haplotype frequencies given the genotypes frequencies. This approach is also known as the gene-counting method (Smith, 1957).

$$p_t^{(s+1)} = \frac{1}{2} \sum_{j=1}^{F} \sum_{i=1}^{c_j} \delta_{it} . pr(genotype_i)^{(s)} \qquad \text{Equation 4}$$

Where $\delta_{it}$ is an indicator variable which count the number of time haplotype $t$ in genotype $i$. It takes the values of 0, 1 or 2.

To ensure that the estimation finally obtained is the maximum-likelihood estimation several values of departures are required. The estimations obtained are compared and if they are different the estimations leading to the best likelihood are kept.

**3) Methods To Calculate Linkage Disequilibrium Between Markers**

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention ($M_i$, $M_j$) having alleles ($a_i/b_i$) at marker $M_i$ and alleles ($a_j/b_j$) at marker $M_j$ can be calculated for every allele combination ($a_i,a_j$; $a_i,b_j$; $b_i,a_j$ and $b_i,b_j$), according to the Piazza formula:

$\Delta_{aiaj} = \sqrt{\theta 4} - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)}$, where:

$\theta 4$= - - = frequency of genotypes not having allele $a_i$ at $M_i$ and not having allele $a_j$ at $M_j$

$\theta 3$= - + = frequency of genotypes not having allele $a_i$ at $M_i$ and having allele $a_j$ at $M_j$

$\theta 2$= + - = frequency of genotypes having allele $a_i$ at $M_i$ and not having allele $a_j$ at $M_j$

Linkage disequilibrium (LD) between pairs of biallelic markers ($M_i$, $M_j$) can also be calculated for every allele combination (ai,aj; ai,bj; $b_i,a_j$ and $b_i,b_j$), according to the maximum-likelihood estimate (MLE) for delta (the composite genotypic disequilibrium coefficient), as described by Weir (Weir B. S., 1996). The MLE for the composite linkage disequilibrium is:

$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4/2)/N - 2(pr(a_i).\ pr(a_j))$

Where $n_1 = \Sigma$ phenotype ($a_i/a_i$, $a_j/a_j$), $n_2 = \Sigma$ phenotype ($a_i/a_i$, $a_j/b_j$), $n_3 = \Sigma$ phenotype ($a_i/b_i$, $a_j/a_j$), n4= $\Sigma$ phenotype ($a_i/b_i$, $a_j/b_j$) and N is the number of individuals in the sample.

This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, $M_i$ ($a_i/b_i$) and $M_j$ ($a_j/b_j$), fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between *ai* and *aj* is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where $pr(a_i)$ is the probability of allele $a_i$ and pr($a_j$) is the probability of allele $a_j$ and where $pr(haplotype\ (a_i, a_j))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between $M_i$ and $M_j$.

Then a normalized value of the above is calculated as follows:

$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i).\ pr(a_j),\ -pr(b_i).\ pr(b_j))$ with $D_{aiaj}<0$

$D'_{aiaj} = D_{aiaj} / \max(pr(b_i).\ pr(a_j),\ pr(a_i).\ pr(b_j))$ with $D_{aiaj}>0$

The skilled person will readily appreciate that other linkage disequilibrium calculation methods can be used.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 75 and 200, more preferably around 100.

**4) Testing For Association**

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and control populations, and comparing these frequencies with a statistical test to determine if their is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype may be used. Preferably the statistical test employed is a chi-square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

Statistical Significance

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about $1 \times 10^{-2}$ or less, more preferably about $1 \times 10^{-4}$ or less, for a single biallelic marker analysis and about $1 \times 10^{-3}$ or less, still more preferably $1 \times 10^{-6}$ or less and most preferably of about $1 \times 10^{-8}$ or less, for a haplotype analysis involving two or more markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and a trait can be revealed and used for diagnosis and drug screening purposes.

Phenotypic Permutation

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control individuals are pooled and randomized with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups, which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is reiterated preferably at least between 100 and 10000 times. The repeated iterations allow the determination of the probability to obtain the tested haplotype by chance.

Assessment Of Statistical Association

To address the problem of false positives similar analysis may be performed with the same case-control populations in random genomic regions. Results in random regions and the candidate region are compared as described in a co-pending US Provisional Patent Application entitled "Methods, Software And Apparati For Identifying Genomic Regions Harboring A Gene Associated With A Detectable Trait," U.S. Serial Number 60/107,986, filed November 10, 1998, the contents of which are incorporated herein by reference.

**5) Evaluation Of Risk Factors**

The association between a risk factor (in genetic epidemiology the risk factor is the presence or the absence of a certain allele or haplotype at marker loci) and a disease is measured by the odds ratio (OR) and by the relative risk (RR). If $P(R^+)$ is the probability of developing the disease for individuals with R and $P(R^-)$ is the probability for individuals without the risk factor, then the relative risk is simply the ratio of the two probabilities, that is:

$$RR = P(R^+)/P(R^-)$$

In case-control studies, direct measures of the relative risk cannot be obtained because of the sampling design. However, the odds ratio allows a good approximation of the relative risk for low-incidence diseases and can be calculated:

$$OR = \left[\frac{F^+}{1-F^+}\right] \Big/ \left[\frac{F^-}{(1-F^-)}\right]$$

$$OR = (F^+/(1-F^+))/(F^-/(1-F^-))$$

$F^+$ is the frequency of the exposure to the risk factor in cases and $F^-$ is the frequency of the exposure to the risk factor in controls. $F^+$ and $F^-$ are calculated using the allelic or haplotype frequencies of the study and further depend on the underlying genetic model (dominant, recessive, additive...).

One can further estimate the attributable risk (AR) which describes the proportion of individuals in a population exhibiting a trait due to a given risk factor. This measure is important in

quantifying the role of a specific factor in disease etiology and in terms of the public health impact of a risk factor. The public health relevance of this measure lies in estimating the proportion of cases of disease in the population that could be prevented if the exposure of interest were absent. AR is determined as follows:

$$AR = P_E (RR-1) / (P_E (RR-1)+1)$$

AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype. $P_E$ is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which, is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

**Identification Of Biallelic Markers In Linkage Disequilibrium With The Biallelic Markers of the Invention**

Once a first biallelic marker has been identified in a genomic region of interest, the practitioner of ordinary skill in the art, using the teachings of the present invention, can easily identify additional biallelic markers in linkage disequilibrium with this first marker. As mentioned before any marker in linkage disequilibrium with a first marker associated with a trait will be associated with the trait. Therefore, once an association has been demonstrated between a given biallelic marker and a trait, the discovery of additional biallelic markers associated with this trait is of great interest in order to increase the density of biallelic markers in this particular region. The causal gene or mutation will be found in the vicinity of the marker or set of markers showing the highest correlation with the trait.

Identification of additional markers in linkage disequilibrium with a given marker involves: (a) amplifying a genomic fragment comprising a first biallelic marker from a plurality of individuals; (b) identifying of second biallelic markers in the genomic region harboring said first biallelic marker; (c) conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and (d) selecting said second biallelic markers as being in linkage disequilibrium with said first marker. Subcombinations comprising steps (b) and (c) are also contemplated.

Methods to identify biallelic markers and to conduct linkage disequilibrium analysis are described herein and can be carried out by the skilled person without undue experimentation. The present invention then also concerns biallelic markers which are in linkage disequilibrium with the specific biallelic markers A1 to A32 and which are expected to present similar characteristics in terms of their respective association with a given trait. In a preferred embodiment, the invention concerns biallelic markers which are in linkage disequilibrium with the specific biallelic markers

**Identification Of Functional Mutations**

Mutations in the *HKLP* gene which are responsible for a detectable phenotype or trait may be identified by comparing the sequences of the *HKLP* gene from trait positive and control

individuals. Once a positive association is confirmed with a biallelic marker of the present invention, the identified locus can be scanned for mutations. In a preferred embodiment, functional regions such as exons and splice sites, promoters and other regulatory regions of the *HKLP* gene are scanned for mutations. In a preferred embodiment the sequence of the *HKLP* gene is compared in trait positive and control individuals. Preferably, trait positive individuals carry the haplotype shown to be associated with the trait and trait negative individuals do not carry the haplotype or allele associated with the trait. The detectable trait or phenotype may comprise a variety of manifestations of altered *HKLP* function.

The mutation detection procedure is essentially similar to that used for biallelic marker identification. The method used to detect such mutations generally comprises the following steps:

- amplification of a region of the *HKLP* gene comprising a biallelic marker or a group of biallelic markers associated with the trait from DNA samples of trait positive patients and trait-negative controls;
- sequencing of the amplified region;
- comparison of DNA sequences from trait positive and control individuals;
- determination of mutations specific to trait-positive patients.

In one embodiment, said biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof. It is preferred that candidate polymorphisms be then verified by screening a larger population of cases and controls by means of any genotyping procedure such as those described herein, preferably using a microsequencing technique in an individual test format. Polymorphisms are considered as candidate mutations when present in cases and controls at frequencies compatible with the expected association results. Polymorphisms are considered as candidate "trait-causing" mutations when they exhibit a statistically significant correlation with the detectable phenotype.

## Recombinant Vectors

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The present invention encompasses a family of recombinant vectors that comprise a regulatory polynucleotide derived from the *HKLP* genomic sequence, and/or a coding polynucleotide from either the *HKLP* genomic sequence or the cDNA sequence.

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, coding sequences and polynucleotide constructs, as well as any *HKLP* primer or probe as defined above. More particularly, the recombinant vectors of the present invention can comprise any of the polynucleotides described in the "Genomic

Sequences Of tThe *HKLP* Gene" section, the "*HKLP* cDNA Sequences" section, the "Coding Regions" section, the "Polynucleotide constructs" section, and the "Oligonucleotide Probes And Primers" section.

In a first preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from a *HKLP* genomic sequence of SEQ ID Nos 1 and 2 or a *HKLP* cDNA, for example the cDNA of SEQ ID No 3 in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

A second preferred embodiment of the recombinant vectors according to the invention consists of expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express the HKLP polypeptide which can be then purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against the HKLP protein. In other embodiments, the expression vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding a HKLP protein, preferably the HKLP protein of the amino acid sequence of SEQ ID No 4 or variants or fragments thereof.

The invention also pertains to a recombinant expression vector useful for the expression of the *HKLP* coding sequence, wherein said vector comprises a nucleic acid of SEQ ID No 3.

Recombinant vectors comprising a nucleic acid containing a *HKLP*-related biallelic marker is also part of the invention. In a preferred embodiment, said biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

**1. General features of the expression vectors of the invention**

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may consist of a chromosomal, non-chromosomal, semi-synthetic and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of:

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp in length that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, said structural or coding sequence being operably linked to the regulatory elements described in (1); and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, when a recombinant protein is expressed without a leader or transport sequence, it may include a N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of the translated protein into the periplasmic space or the extracellular medium. In a specific embodiment wherein the vector is adapted for transfecting and expressing desired sequences in mammalian host cells, preferred vectors will comprise an origin of replication in the desired host, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5'-flanking non-transcribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation sites may be used to provide the required non-transcribed genetic elements.

The *in vivo* expression of a HKLP polypeptide of SEQ ID No 4 or fragments or variants thereof may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of a biologically inactive HKLP protein.

Consequently, the present invention also deals with recombinant expression vectors mainly designed for the *in vivo* production of the HKLP polypeptide of SEQ ID No 4 or fragments or variants thereof by the introduction of the appropriate genetic material in the organism of the patient to be treated. This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

**2. Regulatory Elements**

Promoters

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed. The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell, such as, for example, a human or a viral promoter.

A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the gpt, lambda PR, PL and trp promoters (EP 0036776), the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983; O'Reilly et al., 1992), the lambda PR promoter or also the trc promoter.

Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a promoter is well within the ability of a person skilled in the field of genetic egineering. For example, one may refer to the book of Sambrook et al.(1989) or also to the procedures described by Fuller et al.(1996).

Other regulatory elements

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

**3. Selectable Markers**

Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression construct. The selectable marker genes for selection of

transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

**4. Preferred Vectors.**

Bacterial vectors

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of other suitable vectors are known to those of skill in the art, and commercially available, such as the following bacterial vectors: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

Bacteriophage vectors

The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb.

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994). Recombinant P1 clones comprising *HKLP* nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (Linton et al., 1993). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al.(1994). Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

When the goal is to express a P1 clone comprising *HKLP* nucleotide sequences in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (*Sfi*I, *Not*I or *Sal*I). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar using methods similar to those originally reported for the

isolation of DNA from YACs (Schedl et al., 1993a; Peterson et al., 1993). At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA – 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 µM EDTA) containing 100 mM NaCl, 30 µM spermine, 70 µM spermidine on a microdyalisis membrane (type VS, 0.025 µM from Millipore). The intactness of the purified P1 DNA insert is assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

Baculovirus vectors

A suitable vector for the expression of the HKLP polypeptide of SEQ ID No 3 or fragments or variants thereof is a baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharmingen) that is used to transfect the SF9 cell line (ATCC N°CRL 1711) which is derived from *Spodoptera frugiperda.*

Other suitable vectors for the expression of the HKLP polypeptide of SEQ ID No 3 or fragments or variants thereof in a baculovirus expression system include those described by Chai et al.(1993), Vlasak et al.(1983) and Lenhard et al.(1996).

Viral vectors

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996) or Ohno et al.(1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin ( French patent application N° FR-93.05954).

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides *in vivo* , particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Other preferred retroviral vectors are those described in Roth et al.(1996), PCT Application No WO 93/25234, PCT Application No WO 94/ 06920, Roux et al., 1989, Julan et al., 1992 and Neda et al., 1991.

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

BAC vectors

The bacterial artificial chromosome (BAC) cloning system (Shizuya et al., 1992) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli*. A preferred BAC vector consists of pBeloBAC11 vector that has been described by Kim et al.(1996). BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

**5. Delivery Of The Recombinant Vectors**

In order to effect expression of the polynucleotides and polynucleotide constructs of the invention, these constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states.

One mechanism is viral infection where the expression construct is encapsulated in an infectious viral particle.

Several non-viral methods for the transfer of polynucleotides into cultured mammalian cells are also contemplated by the present invention, and include, without being limited to, calcium phosphate precipitation (Graham et al., 1973; Chen et al., 1987;), DEAE-dextran (Gopal, 1985), electroporation (Tur-Kaspa et al., 1986; Potter et al., 1984), direct microinjection (Harland et al., 1985), DNA-loaded liposomes (Nicolau et al., 1982; Fraley et al., 1979), and receptor-mediated

transfection (Wu and Wu, 1987; 1988). Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application N° WO 90/11092 (Vical Inc.) and also in PCT application No. WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson et al.(1996) and of Huygen et al.(1996).

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be proceeded with a particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein et al.(1987).

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, 1991; Wong et al., 1980; Nicolau et al., 1987)

In a specific embodiment, the invention provides a composition for the *in vivo* production of the HKLP protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 μg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced *in vitro* in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired HKLP polypeptide or the desired fragment

thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

## Cell Hosts

Another object of the invention consists of a host cell that has been transformed or transfected with one of the polynucleotides described herein, and in particular a polynucleotide either comprising a *HKLP* regulatory polynucleotide or the coding sequence of the HKLP polypeptide selected from the group consisting of SEQ ID Nos 1-3 or a fragment or a variant thereof. Also included are host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above. More particularly, the cell hosts of the present invention can comprise any of the polynucleotides described in the "Genomic Sequences Of tThe *HKLP* Gene" section, the "*HKLP* cDNA Sequences" section, the "Coding Regions" section, the "Polynucleotide constructs" section, the "Oligonucleotide Probes And Primers" section and the "Recombinant Vectors" section.

A further recombinant cell host according to the invention comprises a polynucleotide containing a biallelic marker selected from the group consisting of A1 to A32, and the complements thereof.

Preferred host cells used as recipients for the expression vectors of the invention are the following:

a) Prokaryotic host cells: *Escherichia coli* strains (I.E.DH5-α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*.

b) Eukaryotic host cells: HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2), Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711), C127 cells (ATCC N° CRL-1804), 3T3 (ATCC N° CRL-6361), CHO (ATCC N° CCL-61), human kidney 293. (ATCC N° 45504; N° CRL-1573) and BHK (ECACC N° 84100501; N° 84111301).

c) Other mammalian host cells.

The *HKLP* gene expression in mammalian, and typically human, cells may be rendered defective, or alternatively it may be proceeded with the insertion of a *HKLP* genomic or cDNA sequence with the replacement of the *HKLP* gene counterpart in the genome of an animal cell by a *HKLP* polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of cell hosts that may be used are mammal zygotes, such as murine zygotes. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml –for BAC inserts- 3 ng/µl –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250 µM

EDTA containing 100 mM NaCl, 30 μM spermine, and70 μM spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl et al (1993b).

Anyone of the polynucleotides of the invention, including the DNA constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC n° CRL-1821), ES-D3 (ATCC n° CRL1934 and n° CRL-11632), YS001 (ATCC n° CRL-11776), 36.5 (ATCC n° CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells consist of primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo et al.(1993) and are inhibited in growth by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

## Transgenic Animals

The terms "transgenic animals" or "host animals" are used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a HKLP gene disrupted by homologous recombination with a knock out vector.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic

acids comprising a *HKLP* coding sequence, a *HKLP* regulatory polynucleotide, a polynucleotide construct, or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Generally, a transgenic animal according the present invention comprises any one of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention. More particularly, the transgenic animals of the present invention can comprise any of the polynucleotides described in the "Genomic Sequences Of tThe *HKLP* Gene" section, the " *HKLP* cDNA Sequences" section, the "Coding Regions" section, the "Polynucleotide constructs" section, the "Oligonucleotide Probes And Primers" section, the "Recombinant Vectors" section and the "Cell Hosts" section.

A further transgenic animals according to the invention contains in their somatic cells and/or in their germ line cells a polynucleotide comprising a biallelic marker selected from the group consisting of A1 to A32, and the complements thereof.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native HKLP protein, or alternatively a mutant HKLP protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the *HKLP* gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to US Patents Nos 4,873,191, issued Oct. 10, 1989; 5,464,764 issued Nov 7, 1995; and 5,789,215, issued Aug 4, 1998; these documents being herein incorporated by reference to disclose methods producing transgenic mice.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a *HKLP* coding sequence, a *HKLP* regulatory polynucleotide or a DNA sequence encoding a *HKLP* antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas et al.(1987). The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous

recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al.(1988).

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987). The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood et al.(1993) or by Nagy et al.(1993), the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

**Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention.**

A further object of the invention consists of recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or a *HKLP* gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay et al.(1991).

**Methods for screening substances interacting with a HKLP polypeptide**

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the HKLP protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for HKLP or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the HKLP protein is brought into contact with the corresponding purified HKLP protein, for example the corresponding purified recombinant HKLP protein produced by a recombinant cell host as described hereinbefore, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of the HKLP protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No

4, with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang et al. (1997) or the affinity capillary electrophoresis method described by Bush et al. (1997), the disclosures of which are incorporated by reference, can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with the HKLP protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4 may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent , radioactive, or enzymatic tag and placed in contact with immobilized HKLP protein, or a fragment thereof under conditions which permit specific binding to occur. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

Another object of the present invention consists of methods and kits for the screening of candidate substances that interact with HKLP polypeptide.

The present invention pertains to methods for screening substances of interest that interact with a HKLP protein or one fragment or variant thereof. By their capacity to bind covalently or non-covalently to a HKLP protein or to a fragment or variant thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

*In vitro*, said interacting molecules may be used as detection means in order to identify the presence of a HKLP protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance comprises the following steps :

a) providing a polypeptide consisting of a HKLP protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4 or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

The invention further concerns a kit for the screening of a candidate substance interacting with the HKLP polypeptide, wherein said kit comprises :

a) a HKLP protein having an amino acid sequence selected from the group consisting of the amino acid sequences of SEQ ID No 4 or a peptide fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4 or a variant thereof ;

b) optionally means useful to detect the complex formed between the HKLP protein or a peptide fragment or a variant thereof and the candidate substance.

In a preferred embodiment of the kit described above, the detection means consist in monoclonal or polyclonal antibodies directed against the HKLP protein or a peptide fragment or a variant thereof.

Various candidate substances or molecules can be assayed for interaction with a HKLP polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule consists of a polypeptide, this polypeptide may be the resulting expression product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a HKLP polypeptide or a fragment or a variant thereof, and optionally means useful to detect the complex formed between the HKLP polypeptide or its fragment or variant and the candidate substance. In a preferred embodiment the detection means consist in monoclonal or polyclonal antibodies directed against the corresponding HKLP polypeptide or a fragment or a variant thereof.

**A. Candidate ligands obtained from random peptide libraries**

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg K.R. et al., 1992; Valadon P., et al., 1996; Lucas A.H., 1994; Westerink M.A.J., 1995; Felici F. et al., 1991). According to this particular embodiment, the recombinant phages expressing a protein that binds to the immobilized HKLP protein is retained and the complex formed between the HKLP protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the HKLP protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized HKLP protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the HKLP protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti-HKLP, and this phage population is subsequently amplified by an over-infection of bacteria (for example E. coli). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step consists in characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the

phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

**B. Candidate ligands obtained by competition experiments.**

Alternatively, peptides, drugs or small molecules which bind to the HKLP protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4, may be identified in competition experiments. In such assays, the HKLP protein, or a fragment thereof, is immobilized to a surface, such as a plastic plate. Increasing amounts of the peptides, drugs or small molecules are placed in contact with the immobilized HKLP protein, or a fragment thereof, in the presence of a detectable labeled known HKLP protein ligand. For example, the HKLP ligand may be detectably labeled with a fluorescent, radioactive, or enzymatic tag. The ability of the test molecule to bind the HKLP protein, or a fragment thereof, is determined by measuring the amount of detectably labeled known ligand bound in the presence of the test molecule. A decrease in the amount of known ligand bound to the HKLP protein, or a fragment thereof, when the test molecule is present indicated that the test molecule is able to bind to the HKLP protein, or a fragment thereof.

**C. Candidate ligands obtained by affinity chromatography.**

Proteins or other molecules interacting with the HKLP protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4, can also be found using affinity columns which contain the HKLP protein, or a fragment thereof. The HKLP protein, or a fragment thereof, may be attached to the column using conventional techniques including chemical coupling to a suitable column matrix such as agarose, Affi Gel® , or other matrices familiar to those of skill in art. In some embodiments of this method, the affinity column contains chimeric proteins in which the HKLP protein, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins or other molecules interacting with the HKLP protein, or a fragment thereof, attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen et al. (1997), the disclosure of which is incorporated by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

**D. Candidate ligands obtained by optical biosens r meth ds**

Proteins interacting with the HKLP protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4, can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo et al. (1995), the disclosure of which is incorporated by reference. This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules. This technique is based on the surface plasmon resonance (SPR) phenomenon. Briefly, the candidate ligand molecule to be tested is attached to a surface (such as a carboxymethyl dextran matrix). A light beam is directed towards the side of the surface that does not contain the sample to be tested and is reflected by said surface. The SPR phenomenon causes a decrease in the intensity of the reflected light with a specific association of angle and wavelength. The binding of candidate ligand molecules cause a change in the refraction index on the surface, which change is detected as a change in the SPR signal. For screening of candidate ligand molecules or substances that are able to interact with the HKLP protein, or a fragment thereof, the HKLP protein, or a fragment thereof, is immobilized onto a surface. This surface consists of one side of a cell through which flows the candidate molecule to be assayed. The binding of the candidate molecule on the HKLP protein, or a fragment thereof, is detected as a change of the SPR signal. The candidate molecules tested may be proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry. This technique may also be performed by immobilizing eukaryotic or prokaryotic cells or lipid vesicles exhibiting an endogenous or a recombinantly expressed HKLP protein at their surface.

The main advantage of the method is that it allows the determination of the association rate between the HKLP protein and molecules interacting with the HKLP protein. It is thus possible to select specifically ligand molecules interacting with the HKLP protein, or a fragment thereof, through strong or conversely weak association constants.

**E. Candidate ligands obtained through a two-hybrid screening assay.**

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent N° US 5,667,973 and the US Patent N° 5,283,173 (Fields et al.) the technical teachings of both patents being herein incorporated by reference.

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper et al. (1993) or as described by Cho et al. (1998) or also Fromont-Racine et al. (1997).

The bait protein or polypeptide consists of a HKLP polypeptide or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4 or a variant thereof.

More precisely, the nucleotide sequence encoding the HKLP polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "pray" polypeptides.

A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the two different yeast strains may be the followings :

- Y190, the phenotype of which is (*MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh*$^r$);
- Y187, the phenotype of which is (*MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet*$^-$), which is the opposite mating type of Y190.

Briefly, 20 µg of pAS2/*HKLP* and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (*His*$^+$, *beta-gal*$^-$) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/*HKLP* plasmids bu retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing HKLP or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as *Gal4* fusions as described by Harper et al. (1993) and by Bram et al. (1993), and screened for beta galactosidase by filter lift assay. Yeast clones that are *beta gal-* after mating with the control *Gal4* fusions are considered false positives.

In another embodiment of the two-hybrid method according to the invention, interaction between the HKLP or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual

accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), the disclosure of which is incorporated herein by reference, nucleic acids encoding the HKLP protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between HKLP and the protein or peptide encoded by the initially selected cDNA insert.

## Methods For Screening Substances Modulating The Activity Of The HKLP protein

The invention also concerns a method for screening new agents, or candidate substances which modulate the activity of the HKLP protein or a fragment thereof. Preferably, the HKLP protein or a fragment thereof is a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4. Preferably, the candidate substance is mixed with the HKLP protein and the activity of the HKLP protein is measured. Candidate substances include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides.

Various assays for biological activity of motor proteins are known (Sato-Yoshitake et al, 1992 and Scholey, 1993). In vitro motility assays to characterize specific KLPs, for example, include microtubule gliding assays demonstrating translocation of microtubules, organelle movement assays to visualize the movement of a cargo of interest, and single molecule motility assays (Howard 1989, Block 1990) to characterize structural elements.

In short, microtubule gliding assays can be performed by applying a sample containing the HKLP protein to a glass surface without any treatment and incubating with microtubules reconstituted from polymerized tubulin. Microtubule translocation activity and the direction of movement can be determined as in Nangaku 1994 by observing the movement of axonemes on the glass surface. Organelle movement assays can be performed by applying a composition containing the cargo of interest with a solution containing vesicles and the HKLP protein to a glass surface. Movement of the organelle can be observed, for example, by using a cargo-specific fluorescent probe to stain vesicles before incubation with the HKLP protein.

## Methods For Inhibiting The Expression Of A *HKLP* Gen

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of *HKLP* as an antisense tool or a triple helix

tool that inhibits the expression of the corresponding *HKLP* gene. A preferred fragment of the nucleic sequence of *HKLP* comprises an allele of at least one of the biallelic markers A1 to A32.

**Antisense Approach**

Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995).

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the *HKLP* mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of *HKLP* that contains either the translation initiation codon ATG or a splicing donor or acceptor site.

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the *HKLP* mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., (1986) and Izant and Weintraub, (1984), the disclosures of which are incorporated herein by reference.

In some strategies, antisense molecules are obtained by reversing the orientation of the *HKLP* coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of *HKLP* antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi et al.(1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European Patent Application No. EP 0 572 287 A2

An alternative to the antisense technology that is used according to the present invention consists in using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme consists of (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense

ribozymes according to the present invention are prepared as described by Sczakiel et al.(1995), the specific preparation procedures being referred to in said article being herein incorporated by reference.

**Triple Helix Approach**

The *HKLP* genomic DNA may also be used to inhibit the expression of the *HKLP* gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene.

Similarly, a portion of the *HKLP* genomic DNA can be used to study the effect of inhibiting *HKLP* transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the *HKLP* genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the *HKLP* genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting *HKLP* expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting HKLP expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the *HKLP* gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced HKLP expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the *HKLP* gene in cells which have been treated with the oligonucleotide.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above in the antisense approach at a dosage calculated based on the in vitro results, as described in antisense approach.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3′ end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al.(1989), which is hereby incorporated by this reference.

**Computer-R lated Embodiments**

As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any one of the following: a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162; b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884; c) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682; d) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, 2 or 3, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 nucleotide positions of any one of the following ranges of nucleotide positions of: (1) SEQ ID No 1: 1-1000, 1001-2000, 2001-3000, 3001-4000, 4001-5000, 5001-6000, 6001-7000, 7001-8000, 8001-9000, 9001-10000, 10001-11000, 11001-12000, 12001-13000, 13001-14000, 14001-15000, 15001-16000, 16001-17000, 17001-18000, 18001-19000, 19001-20000, 20001-21000, 21001-22000, 22001-23000, 23001-24000, 24001-25000, 25001-26000, 26001-27000, 27001-28000, 28001-29000, 29001-30000, 30001-31000, 31001-32000, 32001-33000, 33001-34000, 34001-35000, 35001-36000, 36001-37000, 37001-38000, 38001-39000, 39001-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162; and (2) SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884; e) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises a G at position 7159 of SEQ ID No 1; f) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises a C either at

position 2551 or 4500 of SEQ ID No 2; g) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises a nucleotide selected in the group consisting of a C at position 5487, and a C at position 6265 of SEQ ID No 3; and, j) a nucleotide sequence complementary to any one of the preceding nucleotide sequences.

The "nucleic acid codes of the invention" further encompass nucleotide sequences homologous to: a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162; b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884 c) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682; and d) sequences complementary to all of the preceding sequences. Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention. It will be appreciated that the nucleic acid codes of the invention can be represented in the traditional single character format (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman & Co., New York.) or in any other format or code which records the identity of the nucleotides in a sequence.

As used herein the term "polypeptide codes of the invention" encompass the polypeptide sequences comprising a contiguous span of at least 6, 8, 10, 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the amino acid positions 1-478 of the SEQ ID No 4. It will be appreciated that the polypeptide codes of the invention can be represented in the traditional single character format or three letter format (See the inside back cover of Stryer, Lubert. Biochemistry, 3rd edition. W. H Freeman & Co., New York.) or in any other format or code which records the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 1. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc.

containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of the invention or the polypeptide codes of the invention stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters

that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of the invention, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of the invention.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The

method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes f the invention through the use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the

steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms The method may be implemented by the computer systems described above and the method illustrated in Figure 3. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention.

An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of the invention.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com).

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg et al., U.S. Patent No. 5,436,850 issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, et al., U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini et al., 1997). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly

generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi et al., 1997).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag et al., 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius$^2$.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.),

the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specification referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the sate of the art to which this invention pertains.

## EXAMPLES

### Example 1

### Identification Of Biallelic Markers - DNA Extraction

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0 4 M
- 200 µl SDS 10%
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

## Example 2

### Identification Of Biallelic Markers: Amplification Of Genomic DNA By PCR

The amplification of specific genomic sequences of the DNA samples of example 1 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

| | |
|---|---|
| Final volume | 25 µl |
| DNA | 2 ng/µl |
| $MgCl_2$ | 2 mM |
| dNTP (each) | 200 µM |
| primer (each) | 2.9 ng/µl |
| Ampli Taq Gold DNA polymerase | 0.05 unit/µl |
| PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl) | 1x |

Each pair of first primers was designed using the sequence information of the *HKLP* gene disclosed herein and the OSP software (Hillier & Green, 1991). This first pair of primers was about 20 nucleotides in length and had the sequences disclosed in Table 1 in the columns labeled PU and RP.

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence: TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence: CAGGAAACAGCTATGACC. The primer containing the additional PU 5' sequence is listed in SEQ ID No 9. The primer containing the additional RP 5' sequence is listed in SEQ ID No 10.

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

**Table 1**

| Amplicon | Position range of the amplicon in SEQ ID 1 | | PU Primer name | Position range of amplification primer in SEQ ID No 1 | | RP Primer name | Complementary position range of amplification primer in SEQ ID No 1 | |
|---|---|---|---|---|---|---|---|---|
| 12-809 | 7041 | 7565 | B1 | 7041 | 7060 | C1 | 7545 | 7565 |
| 12-805 | 16004 | 16483 | B2 | 16004 | 16024 | C2 | 16464 | 16483 |
| 12-790 | 26232 | 26705 | B3 | 26232 | 26252 | C3 | 26685 | 26705 |
| 12-791 | 30902 | 31301 | B4 | 30902 | 30922 | C4 | 31282 | 31301 |
| 12-803 | 33476 | 33932 | B5 | 33476 | 33496 | C5 | 33912 | 33932 |
| 99-33040 | 33934 | 34383 | B6 | 33934 | 33953 | C6 | 34364 | 34383 |
| 12-810 | 34918 | 35369 | B7 | 34918 | 34938 | C7 | 35350 | 35369 |
| 12-787 | 45465 | 45993 | B8 | 45465 | 45485 | C8 | 45974 | 45993 |
| 12-793 | 54361 | 54879 | B9 | 54361 | 54381 | C9 | 54861 | 54879 |
| 12-792 | 56365 | 56813 | B10 | 56365 | 56385 | C10 | 56793 | 56813 |
| 99-41009 | 60111 | 60580 | B11 | 60111 | 60130 | C11 | 60562 | 60580 |
| 12-593 | 78815 | 79349 | B12 | 78815 | 78835 | C12 | 79332 | 79349 |
| 12-589 | 79706 | 80167 | B13 | 79706 | 79726 | C13 | 80149 | 80167 |
| 12-785 | 87264 | 87773 | B14 | 87264 | 87284 | C14 | 87753 | 87773 |
| 12-588 | 101161 | 101704 | B15 | 101161 | 101178 | C15 | 101686 | 101704 |
| 12-603 | 104131 | 104578 | B16 | 104131 | 104151 | C16 | 104558 | 104578 |
| 12-586 | 117017 | 117501 | B17 | 117017 | 117036 | C17 | 117481 | 117501 |
| **Amplicon** | **Position range of the amplicon in SEQ ID 2** | | **PU Primer name** | **Position range of amplification primer in SEQ ID No 2** | | **RP Primer name** | **Complementary position range of amplification primer in SEQ ID No 2** | |
| 12-602 | 2203 | 2620 | B18 | 2203 | 2221 | C18 | 2600 | 2620 |
| 12-587 | 4479 | 4878 | B19 | 4479 | 4499 | C19 | 4858 | 4878 |
| 12-596 | 5996 | 6443 | B20 | 5996 | 6015 | C20 | 6423 | 6443 |
| 12-808 | 10079 | 10543 | B21 | 10079 | 10098 | C21 | 10523 | 10543 |
| Amplicon | **Position range of the amplicon** | | **Primer name** | **Position range of amplification primer** | | **Primer name** | **Complementary position range of amplification primer** | |
| 10-265 | 1 | 357 | B22 | 1 | 18 | C22 | 338 | 357 |
| | SEQ ID No 5 | | | SEQ ID No 5 | | | SEQ ID No 5 | |
| 10-266 | 1 | 420 | B23 | 1 | 20 | C23 | 401 | 420 |
| | SEQ ID No 6 | | | SEQ ID No 6 | | | SEQ ID No 6 | |
| 12-592 | 1 | 465 | B24 | 1 | 19 | C24 | 448 | 465 |
| | SEQ ID No 7 | | | SEQ ID No 7 | | | SEQ ID No 7 | |
| 12-783 | 1 | 449 | B25 | 1 | 19 | C25 | 429 | 449 |
| | SEQ ID No 8 | | | SEQ ID No 8 | | | SEQ ID No 8 | |

## Example 3

## Identification Of Biallelic Markers - Sequencing Of Amplified Gen mic DNA And Identificati n Of Polym rphisms

The sequencing of the amplified DNA obtained in example 2 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software (2.1.2 version)).

The sequence data were further evaluated to detect the presence of biallelic markers within the amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

In the 25 fragments of amplification, 32 biallelic markers were detected. The localization of these biallelic markers are as shown in Table 2.

## Example 4

## Validation Of The Polymorphisms Through Microsequencing

The biallelic markers identified in example 3 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each individual DNA sample described in Example 1.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing were about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the primers used in microsequencing are detailed in Table 4.

The microsequencing reaction was performed as follows :

After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20µl final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then removed by ethanol precipitation. Samples were finally resuspended in formamide-EDTA loading buffer and heated for 2 min at 95°C

before being loaded on a polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is categorized classification as homozygous or heterozygous type based on the height ratio.

Table 2

| Amplicon | BM | Marker Name | Localizati n in *HKLP* gene | P lymor-phism | | BM p siti n in SEQ ID | |
|---|---|---|---|---|---|---|---|
| | | | | all1 | all2 | No 1 | No 3 |
| 12-809 | A1 | 12-809-119 | Exon 3 | G | C | 7159 | 471 |
| 12-805 | A2 | 12-805-115 | Intron 5 | A | G | 16369 | |
| 12-790 | A3 | 12-790-396 | Intron 11 | A | G | 26310 | |
| 12-791 | A4 | 12-791-211 | Intron 14 | A | G | 31112 | |
| 12-803 | A5 | 12-803-125 | Intron 14 | A | T | 33808 | |
| 99-33040 | A6 | 99-33040-321 | Intron 14 | C | T | 34255 | |
| 12-810 | A7 | 12-810-77 | Intron 14 | A | G | 35293 | |
| 12-787 | A8 | 12-787-103 | Intron 21 | A | G | 45892 | |
| 12-793 | A9 | 12-793-383 | Intron 21 | G | T | 54497 | |
| 12-792 | A10 | 12-792-233 | Intron 21 | A | G | 56582 | |
| 99-41009 | A11 | 99-41009-244 | Intron 21 | A | G | 60336 | |
| 99-41009 | A12 | 99-41009-111 | Intron 21 | C | T | 60469 | |
| 12-593 | A13 | 12-593-287 | Intron 26 | - | TAAAT | 79063 | |
| 12-593 | A14 | 12-593-174 | Intron 26 | C | T | 79176 | |
| 12-589 | A15 | 12-589-152 | Intron 26 | G | T | 80016 | |
| 12-785 | A16 | 12-785-200 | Intron 30 | C | T | 87463 | |
| 12-785 | A17 | 12-785-393 | Intron 30 | A | G | 87656 | |
| 12-588 | A18 | 12-588-103 | Intron 36 | A | G | 101602 | |
| 12-603 | A19 | 12-603-191 | Intron 37 | C | T | 104391 | |
| 12-586 | A20 | 12-586-414 | Intron 43 | A | G | 117430 | |
| 12-586 | A21 | 12-586-443 | Intron 43 | - | C | 117459 | |
| **Amplicon** | **BM** | **Marker Name** | **Localization in *HKLP* gene** | **Polymor-phism** | | **BM position in SEQ ID** | |
| | | | | all1 | all2 | No 2 | No 3 |
| 12-602 | A22 | 12-602-196 | Intron 46 | C | T | 2397 | |
| 12-602 | A23 | 12-602-350 | Exon 47 | A | C | 2551 | 5487 |
| 12-587 | A24 | 12-587-379 | Exon 48 | A | C | 4500 | 6265 |
| 12-596 | A25 | 12-596-124 | Exon 48 | A | G | 6119 | 7887 |
| 12-808 | A26 | 12-808-52 | 3' regulatory | A | G | 10130 | |
| 12-808 | A27 | 12-808-75 | 3' regulatory | G | C | 10153 | |
| **Amplicon** | **BM** | **Marker Name** | **Localization** | **Polymor-phism** | | **BM position In SEQ ID No** | |
| | | | | all1 | all2 | | |
| 10-265 | A28 | 10-265-178 | intergenic | A | G | 178 | 5 |
| 10-266 | A29 | 10-266-203 | intergenic | C | T | 203 | 6 |
| 12-592 | A30 | 12-592-118 | intergenic | A | T | 118 | 7 |
| 12-783 | A31 | 12-783-421 | intergenic | C | T | 420 | 8 |
| 12-783 | A32 | 12-783-73 | intergenic | G | C | 72 | 8 |

BM refers to "biallelic marker". All1 and all2 refer respectively to allele 1 and allele 2 of the biallelic marker.

Table 3

| BM | Marker Name | Position range f probes in SEQ ID No 1 | | Probes |
|---|---|---|---|---|
| A1 | 12-809-119 | 7147 | 7171 | P1 |
| A2 | 12-805-115 | 16357 | 16381 | P2 |
| A3 | 12-790-396 | 26298 | 26322 | P3 |
| A4 | 12-791-211 | 31100 | 31124 | P4 |
| A5 | 12-803-125 | 33796 | 33820 | P5 |
| A6 | 99-33040-321 | 34243 | 34267 | P6 |
| A7 | 12-810-77 | 35281 | 35305 | P7 |
| A8 | 12-787-103 | 45880 | 45904 | P8 |
| A9 | 12-793-383 | 54485 | 54509 | P9 |
| A10 | 12-792-233 | 56570 | 56594 | P10 |
| A11 | 99-41009-244 | 60324 | 60348 | P11 |
| A12 | 99-41009-111 | 60457 | 60481 | P12 |
| A14 | 12-593-174 | 79164 | 79188 | P13 |
| A15 | 12-589-152 | 80004 | 80028 | P14 |
| A16 | 12-785-200 | 87451 | 87475 | P15 |
| A17 | 12-785-393 | 87644 | 87668 | P16 |
| A18 | 12-588-103 | 101590 | 101614 | P17 |
| A19 | 12-603-191 | 104379 | 104403 | P18 |
| A20 | 12-586-414 | 117418 | 117442 | P19 |
| **BM** | **Marker Name** | **Position range of probes in SEQ ID No 2** | | **Probes** |
| A22 | 12-602-196 | 2385 | 2409 | P20 |
| A23 | 12-602-350 | 2539 | 2563 | P21 |
| A24 | 12-587-379 | 4488 | 4512 | P22 |
| A25 | 12-596-124 | 6107 | 6131 | P23 |
| A26 | 12-808-52 | 10118 | 10142 | P24 |
| A27 | 12-808-75 | 10141 | 10165 | P25 |
| **BM** | **Marker Name** | **Position range of probes** | | **Probes** |
| A28 | 10-265-178 | 166 | 190 | P26 |
| | | In SEQ ID No 5 | | |
| A29 | 10-266-203 | 191 | 215 | P27 |
| | | In SEQ ID No 6 | | |
| A30 | 12-592-118 | 106 | 130 | P28 |
| | | In SEQ ID No 7 | | |
| A31 | 12-783-421 | 408 | 432 | P29 |
| | | In SEQ ID No 8 | | |
| A32 | 12-783-73 | 60 | 84 | P30 |
| | | In SEQ ID No 8 | | |

Table 4

| Marker Name | Biallelic Marker | Mis. 1 | Position range f microsequencing primer in SEQ ID No 1 | | Mis. 2 | Complementary positi n range of microsequencing primer in SEQ ID No 1 | |
|---|---|---|---|---|---|---|---|
| 12-809-119 | A1 | D1 | 7140 | 7158 | E1 | 7160 | 7178 |
| 12-805-115 | A2 | D2 | 16350 | 16368 | E2 | 16370 | 16388 |
| 12-790-396 | A3 | D3 | 26291 | 26309 | E3 | 26311 | 26329 |
| 12-791-211 | A4 | D4 | 31093 | 31111 | E4 | 31113 | 31131 |
| 12-803-125 | A5 | D5 | 33789 | 33807 | E5 | 33809 | 33827 |
| 99-33040-321 | A6 | D6 | 34236 | 34254 | E6 | 34256 | 34274 |
| 12-810-77 | A7 | D7 | 35274 | 35292 | E7 | 35294 | 35312 |
| 12-787-103 | A8 | D8 | 45873 | 45891 | E8 | 45893 | 45911 |
| 12-793-383 | A9 | D9 | 54478 | 54496 | E9 | 54498 | 54516 |
| 12-792-233 | A10 | D10 | 56563 | 56581 | E10 | 56583 | 56601 |
| 99-41009-244 | A11 | D11 | 60317 | 60335 | E11 | 60337 | 60355 |
| 99-41009-111 | A12 | D12 | 60450 | 60468 | E12 | 60470 | 60488 |
| 12-593-174 | A14 | D13 | 79157 | 79175 | E13 | 79177 | 79195 |
| 12-589-152 | A15 | D14 | 79997 | 80015 | E14 | 80017 | 80035 |
| 12-785-200 | A16 | D15 | 87444 | 87462 | E15 | 87464 | 87482 |
| 12-785-393 | A17 | D16 | 87637 | 87655 | E16 | 87657 | 87675 |
| 12-588-103 | A18 | D17 | 101583 | 101601 | E17 | 101603 | 101621 |
| 12-603-191 | A19 | D18 | 104372 | 104390 | E18 | 104392 | 104410 |
| 12-586-414 | A20 | D19 | 117411 | 117429 | E19 | 117431 | 117449 |
| **Marker Name** | **Biallelic Marker** | **Mis. 1** | **Position range of microsequencing primer in SEQ ID No 2** | | **Mis. 2** | **Complementery position renge of microsequencing primer in SEQ ID No 2** | |
| 12-602-196 | A22 | D20 | 2378 | 2396 | E20 | 2398 | 2416 |
| 12-602-350 | A23 | D21 | 2532 | 2550 | E21 | 2552 | 2570 |
| 12-587-379 | A24 | D22 | 4481 | 4499 | E22 | 4501 | 4519 |
| 12-596-124 | A25 | D23 | 6100 | 6118 | E23 | 6120 | 6138 |
| 12-808-52 | A26 | D24 | 10111 | 10129 | E24 | 10131 | 10149 |
| 12-808-75 | A27 | D25 | 10134 | 10152 | E25 | 10154 | 10172 |
| **Marker Name** | **Biallelic Marker** | **Mis. 1** | **Position range of microsequencing primer** | | **Mis. 2** | **Complementery position renge of microsequencing primer** | |
| 10-265-178 | A28 | D26 | 159 | 177 | E26 | 179 | 197 |
| | | | In SEQ ID No 5 | | | In SEQ ID No 5 | |
| 10-266-203 | A29 | D27 | 184 | 202 | E27 | 204 | 222 |
| | | | In SEQ ID No 6 | | | In SEQ ID No 6 | |
| 12-592-118 | A30 | D28 | 99 | 117 | E28 | 119 | 137 |
| | | | In SEQ ID No 7 | | | In SEQ ID No 7 | |
| 12-783-421 | A31 | D29 | 401 | 419 | E29 | 421 | 439 |
| | | | In SEQ ID No 8 | | | In SEQ ID No 8 | |
| 12-783-73 | A32 | D30 | 53 | 71 | E30 | 73 | 91 |
| | | | In SEQ ID No 8 | | | In SEQ ID No 8 | |

Mis 1 and Mis 2 respectively refer to microsequencing primers which hybridized with the non-coding strand of the *HKLP* gene or with the coding strand of the *HKLP* gene.

## Example 5

## Preparation f Antibody Compositions to the HKLP protein

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the HKLP protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the HKLP protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., (1975) or derivative methods thereof. Also see Harlow, E., and D. Lane. 1988.

Briefly, a mouse is repetitively inoculated with a few micrograms of the HKLP protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the HKLP protein or a portion thereof can be prepared by immunizing suitable non-human animal with the HKLP protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for HKLP concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein by the one skilled in the art without departing from the spirit and scope of the invention.

## FREE TEXT OF THE SEQUENCE LISTING

The following free text appears in the accompanying Sequence Listing :

3'regulatory region

polymorphic base

or

complement

probe

deletion of

insertion of

sequencing oligonucleotide Primer

Artificial Sequence

## REFERENCES

The following references cited herein are incorporated herein by reference in their entireties

Abbondanzo SJ et al., 1993, Methods in Enzymology, Academic Press, New York, pp. 803-823 / Ajioka R.S. et al., *Am. J. Hum. Genet.*, 60:1439-1447, 1997 / Altschul et al., 1990, J. Mol. Biol. 215(3):403-410 / Altschul et al., 1993, Nature Genetics 3:266-272 / Altschul et al., 1997, Nuc. Acids Res. 25:3389-3402 / Anton M. et al., 1995, J. Virol., **69** : 4600-4606 / Araki K et al. (1995) *Proc. Natl. Acad. Sci. USA.* 92(1):160-4. / Aszódi et al., Proteins:Structure, Function, and Genetics, Supplement 1:38-42 (1997) / Baubonis W. (1993) *Nucleic Acids Res.* 21(9):2025-9. / Beaucage et al., *Tetrahedron Lett* 1981, **22**: 1859-1862 / Block et al, 1990, Nature 348, 348-352 / Bradley A., 1987, Production and analysis of chimaeric mice. *In*: E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113. / Bram RJ et al., 1993, Mol. Cell Biol., **13** : 4760-4769 / Brown EL, et al., *Methods Enzymol* 1979;**68**:109-151 / Brutlag et al. Comp. App. Biosci. 6:237-245, 1990 / Bush et al., 1997, J. Chromatogr., **777** : 311-328. / Chai H. et al. (1993) *Biotechnol. Appl. Biochem.*18:259-273. / Chee et al. (1996) *Science.* 274:610-614. / Chen et al. (1987) *Mol. Cell. Biol.* 7:2745-2752. / Chen et al. *Proc. Natl. Acad. Sci. USA* 94/20 10756-10761,1997 / Chen and Kwok *Nucleic Acids Research* 25:347-353 1997 / Cho RJ et al., 1998, Proc. Natl. Acad. Sci. USA, **95**(7) : 3752-3757. / Chou J.Y., 1989, Mol. Endocrinol., **3**: 1511-1514. / Clark A.G. (1990) *Mol. Biol. Evol.* 7:111-122. / Compton J. (1991) *Nature.* 350(6313):91-92. / Davis L.G., et al., Basic Methods in Molecular Biology, ed., Elsevier Press, NY, 1986 / Dempster et al., (1977) *J. R. Stat. Soc.*, 39B:1-38. / Eckner R. et al. (1991) *EMBO J.* 10:3513-3522. / Edwards et Leatherbarrow, *Analytical Biochemistry*, **246**, 1-6 (1997) / Engvall, E., Meth. Enzymol. 70:419 (1980) / Excoffier L. and Slatkin M. (1995) *Mol. Biol. Evol.*, 12(5): 921-927. / Feldman and Steg, 1996, Medecine/Sciences, synthese, 12:47-55 / Felici F., 1991, J. Mol. Biol., Vol. 222:301-310 / Fields and Song, 1989, Nature, **340** : 245-246 / Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980) / Flotte et al. (1992) *Am. J. Respir. Cell Mol. Biol.* 7:349-356. / Fodor et al. (1991) *Science* 251:767-777. / Fraley et al. (1979) *Proc. Natl. Acad. Sci. USA.* 76:3348-3352. / Fromont-Racine M. et al., 1997, Nature Genetics, **16**(3) : 277-282. / Fuller S. A. et al. (1996) *Immunology in Current Protocols in Molecular Biology*, Ausubel et al.Eds, John Wiley & Sons, Inc., USA. / Furth P.A. et al. (1994) *Proc. Natl. Acad. Sci USA.* 91:9302-9306. / Geysen H. Mario et al. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002 / Ghosh and Bacchawat, 1991, *Targeting of liposomes to hepatocytes*, IN: *Liver Diseases, Targeted diagnosis and therapy using specific rceptors and ligands*. Wu et al. Eds., Marcel Dekeker, New York, pp. 87-104. / Gonnet et al., 1992, Science 256:1443-1445 / Gopal (1985) *Mol. Cell. Biol.*, 5:1188-1190. / Gossen

M. et al. (1992) *Proc. Natl. Acad. Sci. USA.* 89:5547-5551. / Gossen M. et al. (1995) *Science.* 268:1766-1769. / Graham et al. (1973) *Virology* 52:456-457. / Green et al., *Ann. Rev. Biochem.* 55:569-597 (1986) / Griffin et al. *Science* 245:967-971 (1989) / Grompe, M. (1993) *Nature Genetics.* 5:111-117. / Grompe, M. et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:5855-5892. / Gu H. et al. (1993) *Cell* 73:1155-1164. / Gu H. et al. (1994) *Science* 265:103-106. / Guatelli J C et al. *Proc. Natl. Acad. Sci. USA.* 35:273-286. / Hacia JG, et al., *Nat Genet* 1996;14(4):441-447 / Haff L. A. and Smirnov I. P. (1997) *Genome Research*, 7:378-388. / Hames B.D. and Higgins S.J. (1985) *Nucleic Acid Hybridization: A Practical Approach.* Hames and Higgins Ed., IRL Press, Oxford. / Harju L, et al., *Clin Chem* 1993;39(11Pt 1):2282-2287 / Harland et al. (1985) *J. Cell. Biol.* 101:1094-1095. / Harlow, E., and D. Lane. 1988. Antibodies A Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242 / Harper JW et al., 1993, Cell, 75 : 805-816 / Hawley M.E. et al. (1994) *Am. J. Phys. Anthropol.* 18:104. / Henikoff and Henikoff, 1993, Proteins 17:49-61 / Higgins et al., 1996, Methods Enzymol. 266:383-402 / Hillier L. and Green P. *Methods Appl.*, 1991, 1: 124-8. / Hoess et al. (1986) *Nucleic Acids Res.* 14:2287-2300. / Howard et al (1989) Nature 342, 154-158 / Huang L. et al. (1996) *Cancer Res* 56(5):1137-1141. / Huygen et al. (1996) *Nature Medicine.* 2(8):893-898. / Izant JG, Weintraub H, *Cell* 1984 Apr;36(4):1007-15 / Julan et al. (1992) *J. Gen. Virol.* 73:3251-3255. / Kanegae Y. et al., 1995 *Nucl. Acids Res.* 23:3816-3821. / Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268 / Khoury J. et al., *Fundamentals of Genetic Epidemiology, Oxford University Press, NY*, 1993 / Kim U-J. et al. (1996) *Genomics* 34:213-218. / Klein et al. (1987) *Nature*. 327:70-73. / Kohler, G. and Milstein, C., Nature 256:495 (1975) / Koller et al. (1992) *Annu. Rev. Immunol.* 10:705-730. / Kozal MJ, et al., *Nat Med* 1996;2(7):753-759 / Landegren U. et al. (1998) *Genome Research*, 8:769-776. / Lander and Schork, *Science*, 265, 2037-2048, 1994 / Lange K. (1997) *Mathematical and Statistical Methods for Genetic Analysis*. Springer, New York. / Lenhard T. et al. (1996) *Gene.* 169:187-190. / Linton M.F. et al. (1993) *J. Clin. Invest.* 92:3029-3037. / Liu Z. et al. (1994) *Proc. Natl. Acad. Sci. USA.* 91: 4528-4262. / Livak KJ, Hainer JW, *Hum Mutat* 1994;3(4):379-385 / Livak et al., *Nature Genetics*, 9:341-342, 1995 / Lucas A.H., 1994, In : Development and Clinical Uses of Haempophilus b Conjugate; / Mackey K, et al., 1998, *Mol Biotechnol*, 9(1):1-5 / Mansour S.L. et al. (1988) *Nature.* 336:348-352. / Marshall R. L. et al. (1994) *PCR Methods and Applications.* 4:80-84. / McCormick et al. (1994) *Genet. Anal. Tech. Appl.* 11:158-164. / McLaughlin B.A. et al. (1996) *Am. J. Hum. Genet.* 59:561-569. / Morton N.E., *Am.J. Hum.Genet.*, 7:277-318, 1955 / Muzyczka et al. (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129. / Nada S. et al. (1993) *Cell* 73:1125-1135. / Nagy A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 8424-8428. / Nangaku M. (1994) *Cell* 79, 1209-1220 / Narang SA, et al., *Methods Enzymol* 1979;68:90-98 / Neda et al. (1991) *J. Biol. Chem.* 266:14143-14146. / Newton et al. (1989) *Nucleic Acids Res.* 17:2503-2516. / Nickerson D.A. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927. / Nicolau

C. et al., 1987, Methods Enzymol., **149**:157-76. / Nicolau et al. (1982) *Biochim. Biophys. Acta.* 721:185-190. / Nyren P, Pettersson B, Uhlen M, *Anal Biochem* 1993;**208**(1):171-175 / O'Reilly et al. (1992) *Baculovirus Expression Vectors: A Laboratory Manual.* W. H. Freeman and Co., New York. / Ohno et al. (1994) *Science.* 265:781-784. / Oldenburg K.R. et al., 1992, Proc. Natl. Acad. Sci., **89**:5393-5397. / Orita et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.*86: 2776-2770. / Ott J., *Analysis of Human Genetic Linkage, John Hopkins University Press, Baltimore,* 1991 / Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973) / Parmley and Smith, Gene, 1988, **73**:305-318 / Pastinen et al., *Genome Research* 1997; 7:606-614 / Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85(8):2444-2448 / Pease S. ans William R.S., 1990, Exp. Cell. Res., **190**: 209-211. / Perlin et al. (1994) *Am. J. Hum. Genet.* 55:777-787. / Peterson et al., 1993, Proc. Natl. Acad. Sci. USA, **90** : 7593-7597. / Potter et al. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81(22):7161-7165. / Ramunsen et al., 1997, Electrophoresis, **18** : 588-598. / Reid L.H. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:4299-4303. / Risch, N. and Merikangas, K. (*Science*, 273:1516-1517, 1996 / Robertson E., 1987, Embryo-derived stem cell lines. In: E.J. Robertson Ed. *Teratocarcinomas and embrionic stem cells: a practical approach.* IRL Press, Oxford, pp. 71. / Rossi et al., *Pharmacol. Ther.* **50**:245-254, (1991) / Roth J.A. et al. (1996) *Nature Medicine.* 2(9):985-991. / Roux et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:9079-9083. / Ruano et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:6296-6300. / Sambrook, J., et al. (1989) *Molecular Cloning: A Laboratory Manual.* 2ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. / Samson M, et al. (1996) *Nature*, 382(6593):722-725. / Samulski et al. (1989) *J. Virol.* 63:3822-3828. / Sanchez-Pescador R. (1988) *J. Clin. Microbiol.* 26(10):1934-1938. / Sarkar, G. and Sommer S.S. (1991) *Biotechniques.* / Sato-Yoshitake et al., J Biol Chem 1992 Nov 25;267(33):23930-6 / Sauer B. et al. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:5166-5170. / Schedl A. et al., 1993a, Nature, **362**: 258-261. / Schedl et al., 1993b, Nucleic Acids Res., **21**: 4783-4787. / Schneider et al.(1997) *Arlequin: A Software For Population Genetics Data Analysis.* University of Geneva. / Scholey, 1993, Motility assays for Motor Proteins, (San Diego, California : Academic Press) / Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation / Schaid D.J. et al., *Genet. Epidemiol.*,13:423-450, 1996 / Sczakiel G. et al. (1995) *Trends Microbiol.* 3(6):213-217. / Shay J.W. et al., 1991, Biochem. Biophys. Acta, **1072**: 1-7. / Sheffield, V.C. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 49:699-706. / Shizuya et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797. / Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW, *Nat Genet* 1996;**14**(4):450-456 / Spielmann S. et al., *Am. J. Hum. Genet.*, 52:506-516, 1993 / Spielmann S. and Ewens W.J., *Am. J. Hum. Genet.*, 62:450-458, 1998 / Smith (1957) *Ann. Hum. Genet.* 21:254-276. / Smith et al. (1983) *Mol. Cell. Biol.* 3:2156-2165. / Sowdhamini et al., Protein Engineering 10:207, 215 (1997) / Sternberg N.L. (1992) Trends Genet. 8:1-16. /

Sternberg N.L. (1994) *Mamm. Genome.* 5:397-404. / Syvanen AC, *Clin Chim Acta* 1994;226(2):225-236 / Szabo A. et al. *Curr Opin Struct Biol* 5, 699-705 (1995) / Tacson et al. (1996) *Nature Medicine.* 2(8):888-892. / Te Riele et al. (1990) Nature. 348:649-651. / Terwilliger J.D. and Ott J., *Handbook of Human Genetic Linkage, John Hopkins University Press, London*, 1994 / Thomas K.R. et al. (1986) *Cell.* 44:419-428. / Thomas K.R. et al. (1987) *Cell.* 51:503-512. / Thompson et al., 1994, Nucleic Acids Res. 22(2):4673-4680 / Tur-Kaspa et al. (1986) *Mol. Cell. Biol.* 6:716-718. / Tyagi et al. (1998) *Nature Biotechnology.* 16:49-53. / Urdea M.S. (1988) *Nucleic Acids Research.* 11:4937-4957. / Urdea M.S. et al.(1991) *Nucleic Acids Symp. Ser.* 24:197-200. / Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33:988-991 (1971) / Valadon P., et al., 1996, J. Mol. Biol., **261**:11-22. / Van der Lugt et al. (1991) *Gene.* 105:263-267. / Vlasak R. et al. (1983) *Eur. J. Biochem.* 135:123-126. / Wabiko et al. (1986) *DNA.*5(4):305-314. / Walker et al. (1996) *Clin. Chem.* 42:9-13. / Wang et al., 1997, Chromatographia, 44 : 205-208. / Weir, B.S. (1996) *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, U.S.A.* / Westerink M.A.J., 1995, Proc. Natl. Acad. Sci., **92**:4021-4025 / White, M.B. et al. (1992) *Genomics.* 12:301-306. / White, M.B. et al. (1997) *Genomics.* 12:301-306. / Wong et al. (1980) *Gene.* 10:87-94. / Wood S.A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 4582-4585. / Wu and Wu (1987) *J. Biol. Chem.* 262:4429-4432. / Wu and Wu (1988) *Biochemistry.* 27:887-892. / Wu et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:2757. / Yagi T. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:9918-9922. / Yonekawa Y. et al. (1998) *J. Cell. Biol.* **141**, 431-441 / Zhao et al., *Am. J. Hum. Genet.*, 63:225-240, 1998 / Zou Y. R. et al. (1994) *Curr. Biol.* 4:1099-1103.

## CLAIMS

1. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises either :

- at least 1 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162; or,

- a G at position 7159 of SEQ ID No 1.

2. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 2 or the complementary sequence thereof, wherein said contiguous span comprises:

- at least 1 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884; or,

- a C either at position 2551 or 4500 of SEQ ID No 2.

3. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complementary sequence thereof, wherein said contiguous span comprises either :

- at least 1 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682; or,

- a C at position 5487, or a C at position 6265 of SEQ ID No 3.

4. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID Nos 1-3 and 5-8 or the complement thereof, wherein said span includes a *HKLP*-related biallelic marker in said sequence.

5. A polynucleotide according to claim 4, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof.

6. A polynucleotide according to claim 4, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A22 and A25 to A32, and the complements thereof.

7. A polynucleotide according to claim 4, wherein said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof.

8. A polynucleotide according to any one of claims 4 to 7, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

9. A polynucleotide according to claim 8, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide.

10. A polynucleotide according to claim 9, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: P1 to P30, and the complementary sequences thereto.

11. A polynucleotide according to any one of claims 4 to 7, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

12. A polynucleotide according to any one of claims 1 to 3, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.

13. A polynucleotide according to claim 12, wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of a *HKLP*-related biallelic marker in said sequence.

14. A polynucleotide according to claim 13, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said *HKLP*-related biallelic marker in said sequence.

15. A polynucleotide according to claim 14, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D30, and E1 to E30.

16. An isolated, purified, or recombinant polynucleotide consisting essentially of a sequence selected from the following sequences: B1 to B19 and C1 to C25.

17. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4.

18. A polynucleotide according to any one of claims 1 to 17 attached to a solid support.

19. An array of polynucleotides comprising at least one polynucleotide according to claim 18.

20. An array according to claim 19, wherein said array is addressable.

21. A polynucleotide according to any one of claims 1 to 17 further comprising a label.

22. A recombinant vector comprising a polynucleotide according to any one of claims 1 to 17.

23. A host cell comprising a recombinant vector according to claim 22.

24. A non-human host animal or mammal comprising a recombinant vector according to claim 22.

25. A mammalian host cell comprising an *HKLP* gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1 to 17.

26. A non-human host mammal comprising a *HKLP* gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1 to 17.

27. A method of genotyping comprising determining the identity of a nucleotide at a *HKLP*-related biallelic marker or the complement thereof in a biological sample.

28. A method according to claim 27, wherein said biological sample is derived from a single subject.

29. A method according to claim 28, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome.

30. A method according to claim 27, wherein said biological sample is derived from multiple subjects.

31. A method according to claim 27, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

32. A method according to claim 31, wherein said amplifying is performed by PCR.

33. A method according to claim 27, wherein said determining is performed by a hybridization assay.

34. A method according to claim 27, wherein said determining is performed by a sequencing assay.

35. A method according to claim 27, wherein said determining is performed by a microsequencing assay.

36. A method according to claim 27, wherein said determining is performed by an enzyme-based mismatch detection assay.

37. A method of estimating the frequency of an allele of a *HKLP*-related biallelic marker in a population comprising:

a) genotyping individuals from said population for said biallelic marker according to the method of claim 27; and

b) determining the proportional representation of said biallelic marker in said population.

38. A method of detecting an association between a genotype and a trait, comprising the steps of:

a) determining the frequency of at least one *HKLP*-related biallelic marker in trait positive population according to the method of claim 37;

b) determining the frequency of said *HKLP*-related biallelic marker in a control population according to the method of claim 37; and

c) determining whether a statistically significant association exists between said genotype and said trait.

39. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

a) genotyping at least one *HKLP*-related biallelic marker according to claim 29 for each individual in said population;

b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and

c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

40. A method according to claim 39, wherein said haplotype determination method is selected from the group consisting of asymmetric PCR amplifcation, double PCR amplification of specific alleles, the Clark method, or an expectation maximization algorithm.

41. A method of detecting an association between a haplotype and a trait, comprising the steps of:

a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 39;

b) estimating the frequency of said haplotype in a control population according to the method of claim 39; and

c) determining whether a statistically significant association exists between said haplotype and said trait.

42. A method according to claim 38, wherein said genotyping steps a) and b) are performed on a single pooled biological sample derived from each of said populations.

43. A method according to claim 38, wherein said genotyping steps a) and b) performed separately on biological samples derived from each individual in said populations.

44. A method according to either claim 38 or 41, wherein said control population is a trait negative population.

45. A method according to either claim 38 or 41, wherein said case control population is a random population.

46. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4.

47. An isolated or purified antibody composition are capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 46, wherein said epitope comprises at least 1 of the amino acid positions 1-478 of the SEQ ID No 4.

48. A method according to any one of claims 27 to 45 wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A32, and the complements thereof.

49. A method according to any one of claims 27 to 45 wherein said *HKLP*-related biallelic marker is selected from the group consisting of A1 to A22 and A25 to A32, and the complements thereof.

50. A method according to any one of claims 27 to 45 wherein said *HKLP*-related biallelic marker is selected from the group consisting of A23 and A24, and the complements thereof.

51. A diagnostic kit comprising a polynucleotide according to any one of claims 1 to 21.

52. Use of a polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of the SEQ ID Nos 1-3 and 5-8 or the complementary sequence thereto for determining the identity of the nucleotide at a *HKLP*-related biallelic marker

53. Use according to claim 52 in a microsequencing assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said *HKLP* related biallelic marker in said sequence.

54. Use according to claim 52 in a hybridization assay, wherein said span includes said *HKLP* -related biallelic marker.

55. Use according to claim 52 in a specific amplification assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

56. Use according to claim 52 in a sequencing assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide.

57. Use according to any one of claims 52-56, wherein said *HKLP* -related biallelic is a biallelic marker selected in the group consisting of the biallelic markers A1 to A32.

58. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884;

c) a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises a G at position 7159 of SEQ ID No 1;

d) a contiguous span of at least 12 nucleotides of SEQ ID No 4 or the complements thereof, wherein said contiguous span comprises a C either at position 2551 or 4500 of SEQ ID No 4;

e) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682;

f) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises a nucleotide selected in the group consisting of a C at position 5487, and a C at position 6265 of SEQ ID No 3; and

g) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), e), or f).

59. A computer readable medium having stored thereon a sequence consisting of a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4.

60. A computer system comprising a processor and a data storage device wherein said data storage device a computer readable medium according to with claim 58 or 59.

61. A computer system according to claim 60, further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

62. A computer system of Claim 61 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

63. A computer system of Claim 60 further comprising an identifier which identifies features in said sequence.

64. A method for comparing a first sequence to a reference sequence, comprising the steps of:

reading said first sequence and said reference sequence through use of a computer program which compares sequences; and

determining differences between said first sequence and said reference sequence with said computer program,

wherein said first sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884;

c) a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises a G at position 7159 of SEQ ID No 1;

d) a contiguous span of at least 12 nucleotides of SEQ ID No 4 or the complements thereof, wherein said contiguous span comprises a C either at position 2551 or 4500 of SEQ ID No 4;

e) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682;

f) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises a nucleotide selected in the group consisting of a C at position 5487, and a C at position 6265 of SEQ ID No 3; and

g) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), e), or f); and

a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4.

65. A method according to Claim 64, wherein said step of determining differences between the first sequence and the reference sequence comprises identifying at least one polymorphism.

66. A method for identifying a feature in a sequence, comprising the steps of:

reading said sequence through the use of a computer program which identifies features in sequences; and

identifying features in said sequence with said computer program;

wherein said sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-39624, 39705-40589, 40666-43629, 43710-44203, 44311-45125, 45210-45440, 45622-45717, 45791-68580, 68675-70246, 70396-72421, 72601-73295, 73434-74648, 74898-83055, 83175-85192, 85279-85609, 85740-85906, 86070-88304, 88396-90585, 90705-91767, 91824-94380, 94490-96296, 96364-97184, 97270-101167, 101274-109465, 109581-110228, 110363-111819, 111882-113636, 113783-113945, 114186-117002, 117075-119676, and 119677-121162;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-1600, 1751-2138, 2332-2539, 2659-3829 and 8885-10884;

c) a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises a G at position 7159 of SEQ ID No 1;

d) a contiguous span of at least 12 nucleotides of SEQ ID No 4 or the complements thereof, wherein said contiguous span comprises a C either at position 2551 or 4500 of SEQ ID No 4;

e) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 3: 391-1619 and 6988-10682;

f) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises a nucleotide selected in the group consisting of a C at position 5487, and a C at position 6265 of SEQ ID No 3; and

g) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), e), or f); and

a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4.

67. A method for the screening of a candidate substance interacting with a HKLP polyeptide comprising the following steps :

a) providing a polypeptide consisting of a HKLP protein or a fragment comprising a contiguous span of at least 6 amino acids amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-478 of the SEQ ID No 4 or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

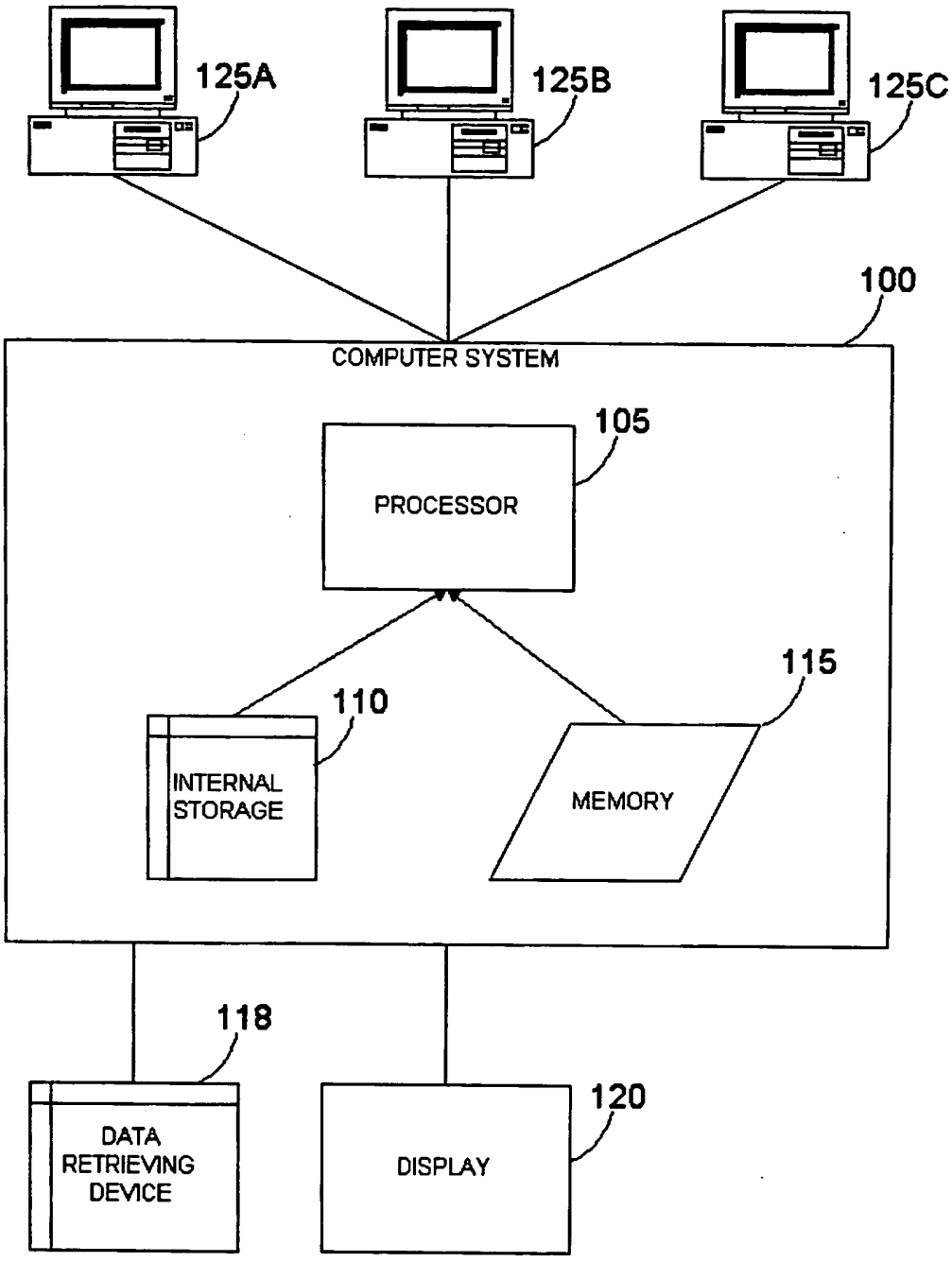
125A
125B
125C
100
COMPUTER SYSTEM
105
PROCESSOR
115
110
INTERNAL STORAGE
MEMORY
118
DATA RETRIEVING DEVICE
120
DISPLAY

FIG.1

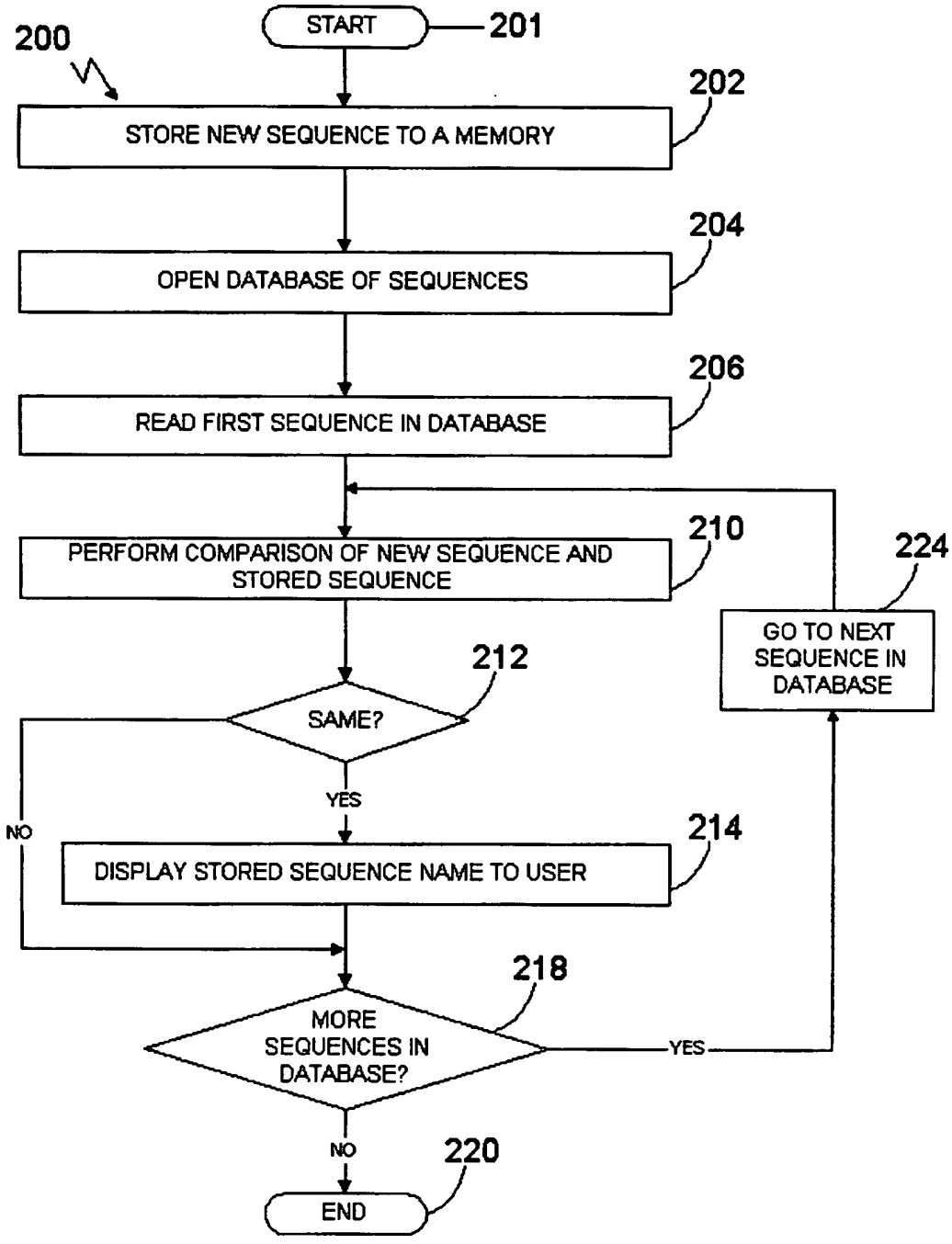
200
START
201
STORE NEW SEQUENCE TO A MEMORY
202
OPEN DATABASE OF SEQUENCES
204
READ FIRST SEQUENCE IN DATABASE
206
PERFORM COMPARISON OF NEW SEQUENCE AND STORED SEQUENCE
210
SAME?
212
YES
NO
DISPLAY STORED SEQUENCE NAME TO USER
214
MORE SEQUENCES IN DATABASE?
218
YES
NO
END
220
GO TO NEXT SEQUENCE IN DATABASE
224

FIG.2

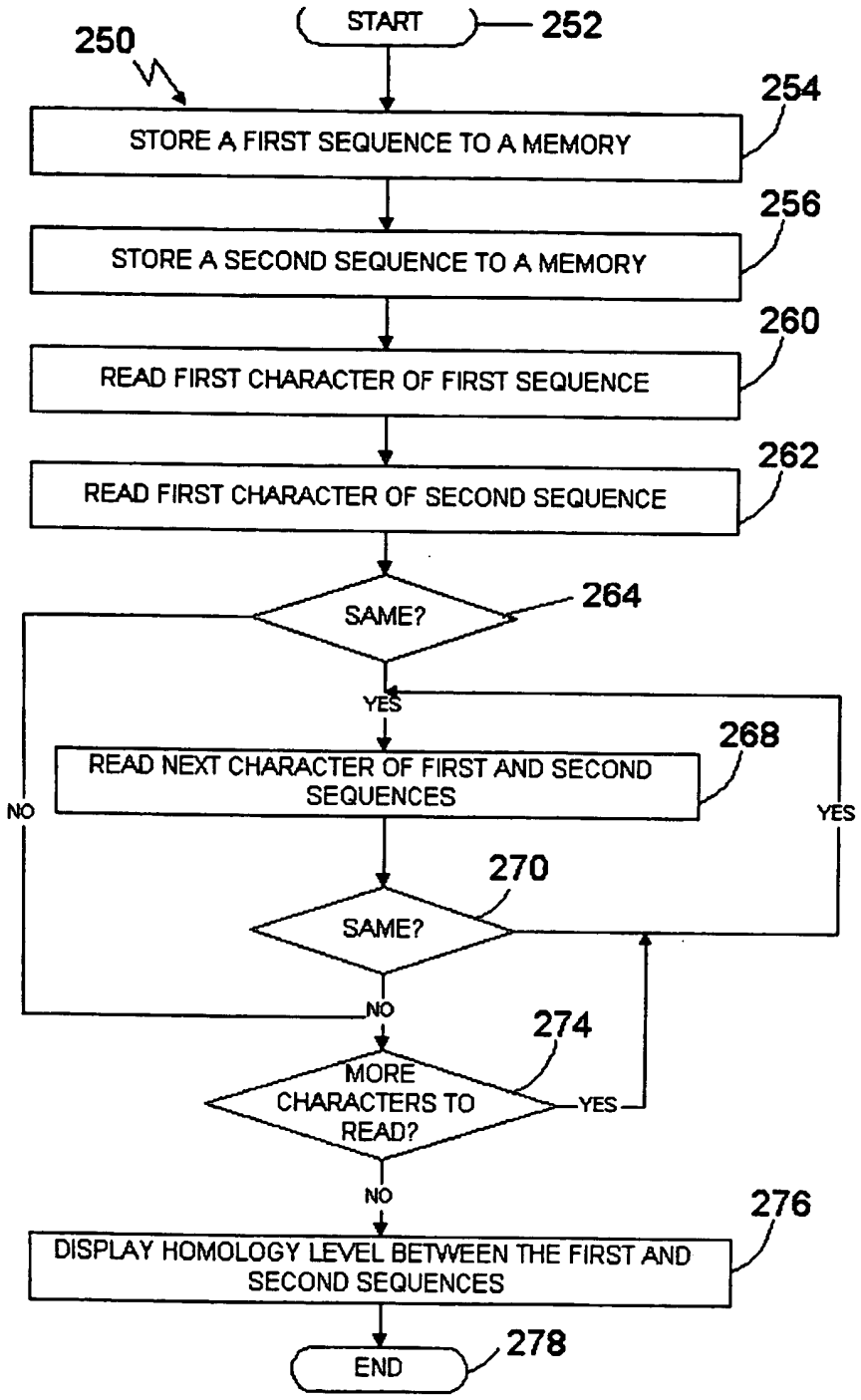
START
252
250
254
STORE A FIRST SEQUENCE TO A MEMORY
256
STORE A SECOND SEQUENCE TO A MEMORY
260
READ FIRST CHARACTER OF FIRST SEQUENCE
262
READ FIRST CHARACTER OF SECOND SEQUENCE
SAME?
264
YES
268
READ NEXT CHARACTER OF FIRST AND SECOND SEQUENCES
NO
YES
270
SAME?
NO
274
MORE CHARACTERS TO READ?
YES
NO
276
DISPLAY HOMOLOGY LEVEL BETWEEN THE FIRST AND SECOND SEQUENCES
END
278

FIG.3

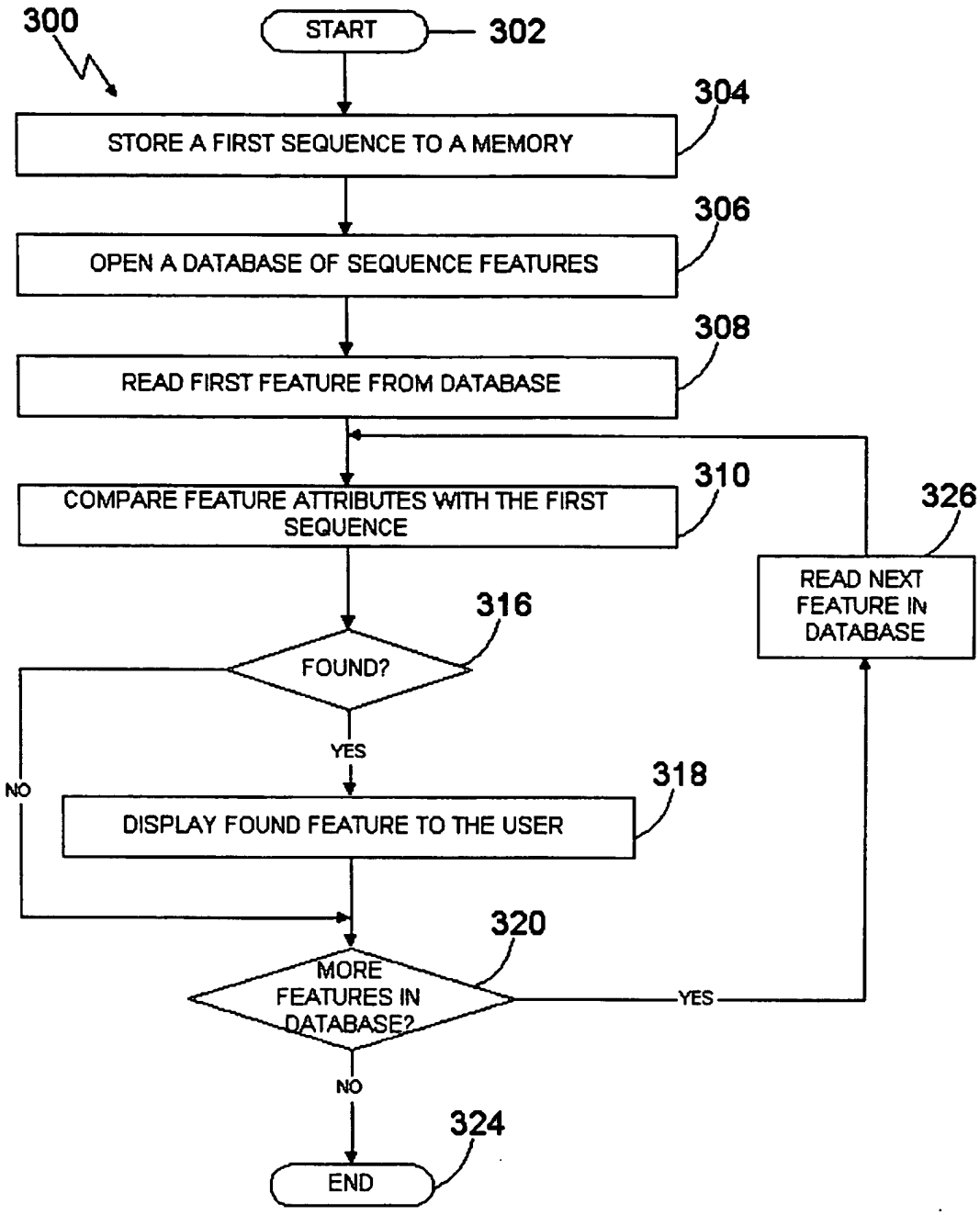
300
START
302
304
STORE A FIRST SEQUENCE TO A MEMORY
306
OPEN A DATABASE OF SEQUENCE FEATURES
308
READ FIRST FEATURE FROM DATABASE
310
COMPARE FEATURE ATTRIBUTES WITH THE FIRST SEQUENCE
326
READ NEXT FEATURE IN DATABASE
316
FOUND?
YES
NO
318
DISPLAY FOUND FEATURE TO THE USER
320
MORE FEATURES IN DATABASE?
YES
NO
324
END

FIG.4

<110> Genset

<120> DNA encoding a novel kinesin-like protein

<130> 64.WO1

<150> US 60/130,217
<151> 1999-04-20

<160> 10

<170> Patent.pm

<210> 1
<211> 121162
<212> DNA
<213> Homo sapiens

<220>
<221> exon
<222> 4844..4920
<223> exon

<220>
<221> exon
<222> 7058..7237
<223> exon

<220>
<221> exon
<222> 10469..10534
<223> exon

<220>
<221> exon
<222> 15940..16118
<223> exon

<220>
<221> exon
<222> 16712..16823
<223> exon

<220>
<221> exon
<222> 20054..20131
<223> exon

<220>
<221> exon
<222> 20793..20858
<223> exon

<220>
<221> exon
<222> 21558..21575
<223> exon

<220>
<221> exon
<222> 23971..24046
<223> exon

<220>
<221> exon
<222> 24864..24942
<223> exon

<220>
<221> exon
<222> 26529..26671
<223> exon

<220>
<221> exon
<222> 27640..27681
<223> exon

<220>
<221> exon
<222> 30863..31074
<223> exon

<220>
<221> exon
<222> 39625..39704
<223> exon

<220>
<221> exon
<222> 40590..40665
<223> exon

<220>
<221> exon
<222> 43630..43709
<223> exon

<220>
<221> exon
<222> 44204..44310
<223> exon

<220>
<221> exon
<222> 45126..45209
<223> exon

<220>
<221> exon
<222> 45441..45621
<223> exon

<220>
<221> exon
<222> 45718..45790
<223> exon

<220>
<221> exon
<222> 68581..68674
<223> exon

<220>
<221> exon

<222> 70247..70395
<223> exon

<220>
<221> exon
<222> 72422..72600
<223> exon

<220>
<221> exon
<222> 73296..73433
<223> exon

<220>
<221> exon
<222> 74649..74897
<223> exon

<220>
<221> exon
<222> 83056..83174
<223> exon

<220>
<221> exon
<222> 85193..85278
<223> exon

<220>
<221> exon
<222> 85610..85739
<223> exon

<220>
<221> exon
<222> 85907..86069
<223> exon

<220>
<221> exon
<222> 88305..88395
<223> exon

<220>
<221> exon
<222> 90586..90704
<223> exon

<220>
<221> exon
<222> 91768..91823
<223> exon

<220>
<221> exon
<222> 94381..94489
<223> exon

<220>
<221> exon
<222> 96297..96363
<223> exon

<220>
<221> exon
<222> 97185..97269
<223> exon

<220>
<221> exon
<222> 101168..101273
<223> exon

<220>
<221> exon
<222> 109466..109580
<223> exon

<220>
<221> exon
<222> 110229..110362
<223> exon

<220>
<221> exon
<222> 111820..111881
<223> exon

<220>
<221> exon
<222> 113637..113782
<223> exon

<220>
<221> exon
<222> 113946..114185
<223> exon

<220>
<221> exon
<222> 117003..117074
<223> exon

<220>
<221> exon
<222> 119677..119798
<223> exon

<220>
<221> allele
<222> 7159
<223> 12-809-119 : polymorphic base G or C

<220>
<221> allele
<222> 16369
<223> 12-805-115 : polymorphic base A or G

<220>
<221> allele
<222> 26310
<223> 12-790-396 : polymorphic base A or G

<220>
<221> allele
<222> 31112

<223> 12-791-211 : polymorphic bas A or G

<220>
<221> allele
<222> 33808
<223> 12-803-125 : polymorphic base A or T

<220>
<221> allele
<222> 34255
<223> 99-33040-321 : polymorphic base C or T

<220>
<221> allele
<222> 35293
<223> 12-810-77 : polymorphic base A or G

<220>
<221> allele
<222> 45892
<223> 12-787-103 : polymorphic base A or G

<220>
<221> allele
<222> 54497
<223> 12-793-383 : polymorphic base G or T

<220>
<221> allele
<222> 56582
<223> 12-792-233 : polymorphic base A or G

<220>
<221> allele
<222> 60336
<223> 99-41009-244 : polymorphic base A or G

<220>
<221> allele
<222> 60469
<223> 99-41009-111 : polymorphic base C or T

<220>
<221> allele
<222> 79063
<223> 12-593-287 : insertion of TAAAT

<220>
<221> allele
<222> 79176
<223> 12-593-174 : polymorphic base C or T

<220>
<221> allele
<222> 80016
<223> 12-589-152 : polymorphic base G or T

<220>
<221> allele
<222> 87463
<223> 12-785-200 : polymorphic base C or T

<220>

<221> allele
<222> 87656
<223> 12-785-393 : polymorphic base A or G

<220>
<221> allele
<222> 101602
<223> 12-588-103 : polymorphic base A or G

<220>
<221> allele
<222> 104391
<223> 12-603-191 : polymorphic base C or T

<220>
<221> allele
<222> 117430
<223> 12-586-414 : polymorphic base A or G

<220>
<221> allele
<222> 117459
<223> 12-586-443 : deletion of C

<220>
<221> primer_bind
<222> 7041..7060
<223> 12-809.pu

<220>
<221> primer_bind
<222> 7545..7565
<223> 12-809.rp complement

<220>
<221> primer_bind
<222> 16004..16024
<223> 12-805.rp

<220>
<221> primer_bind
<222> 16464..16483
<223> 12-805.pu complement

<220>
<221> primer_bind
<222> 26232..26252
<223> 12-790.rp

<220>
<221> primer_bind
<222> 26685..26705
<223> 12-790.pu complement

<220>
<221> primer_bind
<222> 30902..30922
<223> 12-791.pu

<220>
<221> primer_bind
<222> 31282..31301
<223> 12-791.rp complement

<220>
<221> primer_bind
<222> 33476..33496
<223> 12-803.rp

<220>
<221> primer_bind
<222> 33912..33932
<223> 12-803.pu complement

<220>
<221> primer_bind
<222> 33934..33953
<223> 99-33040.pu

<220>
<221> primer_bind
<222> 34364..34383
<223> 99-33040.rp complement

<220>
<221> primer_bind
<222> 34918..34938
<223> 12-810.rp

<220>
<221> primer_bind
<222> 35350..35369
<223> 12-810.pu complement

<220>
<221> primer_bind
<222> 45465..45485
<223> 12-787.rp

<220>
<221> primer_bind
<222> 45974..45993
<223> 12-787.pu complement

<220>
<221> primer_bind
<222> 54361..54381
<223> 12-793.rp

<220>
<221> primer_bind
<222> 54861..54879
<223> 12-793.pu complement

<220>
<221> primer_bind
<222> 56365..56385
<223> 12-792.rp

<220>
<221> primer_bind
<222> 56793..56813
<223> 12-792.pu complement

<220>
<221> primer_bind

<222> 60111..60130
<223> 99-41009.rp

<220>
<221> primer_bind
<222> 60562..60580
<223> 99-41009.pu complement

<220>
<221> primer_bind
<222> 78815..78835
<223> 12-593.rp

<220>
<221> primer_bind
<222> 79332..79349
<223> 12-593.pu complement

<220>
<221> primer_bind
<222> 79706..79726
<223> 12-589.rp

<220>
<221> primer_bind
<222> 80149..80167
<223> 12-589.pu complement

<220>
<221> primer_bind
<222> 87264..87284
<223> 12-785.pu

<220>
<221> primer_bind
<222> 87753..87773
<223> 12-785.rp complement

<220>
<221> primer_bind
<222> 101161..101178
<223> 12-588.rp

<220>
<221> primer_bind
<222> 101686..101704
<223> 12-588.pu complement

<220>
<221> primer_bind
<222> 104131..104151
<223> 12-603.rp

<220>
<221> primer_bind
<222> 104558..104578
<223> 12-603.pu complement

<220>
<221> primer_bind
<222> 117017..117036
<223> 12-586.pu

<220>
<221> primer_bind
<222> 117481..117501
<223> 12-586.rp complement

<220>
<221> primer_bind
<222> 7140..7158
<223> 12-809-119.mis

<220>
<221> primer_bind
<222> 7160..7178
<223> 12-809-119.mis complement

<220>
<221> primer_bind
<222> 16350..16368
<223> 12-805-115.mis

<220>
<221> primer_bind
<222> 16370..16388
<223> 12-805-115.mis complement

<220>
<221> primer_bind
<222> 26291..26309
<223> 12-790-396.mis

<220>
<221> primer_bind
<222> 26311..26329
<223> 12-790-396.mis complement

<220>
<221> primer_bind
<222> 31093..31111
<223> 12-791-211.mis

<220>
<221> primer_bind
<222> 31113..31131
<223> 12-791-211.mis complement

<220>
<221> primer_bind
<222> 33789..33807
<223> 12-803-125.mis

<220>
<221> primer_bind
<222> 33809..33827
<223> 12-803-125.mis complement

<220>
<221> primer_bind
<222> 34236..34254
<223> 99-33040-321.mis

<220>
<221> primer_bind
<222> 34256..34274

<223> 99-33040-321.mis complement

<220>
<221> primer_bind
<222> 35274..35292
<223> 12-810-77.mis

<220>
<221> primer_bind
<222> 35294..35312
<223> 12-810-77.mis complement

<220>
<221> primer_bind
<222> 45873..45891
<223> 12-787-103.mis

<220>
<221> primer_bind
<222> 45893..45911
<223> 12-787-103.mis complement

<220>
<221> primer_bind
<222> 54478..54496
<223> 12-793-383.mis

<220>
<221> primer_bind
<222> 54498..54516
<223> 12-793-383.mis complement

<220>
<221> primer_bind
<222> 56563..56581
<223> 12-792-233.mis

<220>
<221> primer_bind
<222> 56583..56601
<223> 12-792-233.mis complement

<220>
<221> primer_bind
<222> 60317..60335
<223> 99-41009-244.mis

<220>
<221> primer_bind
<222> 60337..60355
<223> 99-41009-244.mis complement

<220>
<221> primer_bind
<222> 60450..60468
<223> 99-41009-111.mis

<220>
<221> primer_bind
<222> 60470..60488
<223> 99-41009-111.mis complement

<220>

<221> primer_bind
<222> 79157..79175
<223> 12-593-174.mis

<220>
<221> primer_bind
<222> 79177..79195
<223> 12-593-174.mis complement

<220>
<221> primer_bind
<222> 79997..80015
<223> 12-589-152.mis

<220>
<221> primer_bind
<222> 80017..80035
<223> 12-589-152.mis complement

<220>
<221> primer_bind
<222> 87444..87462
<223> 12-785-200.mis

<220>
<221> primer_bind
<222> 87464..87482
<223> 12-785-200.mis complement

<220>
<221> primer_bind
<222> 87637..87655
<223> 12-785-393.mis

<220>
<221> primer_bind
<222> 87657..87675
<223> 12-785-393.mis complement

<220>
<221> primer_bind
<222> 101583..101601
<223> 12-588-103.mis

<220>
<221> primer_bind
<222> 101603..101621
<223> 12-588-103.mis complement

<220>
<221> primer_bind
<222> 104372..104390
<223> 12-603-191.mis

<220>
<221> primer_bind
<222> 104392..104410
<223> 12-603-191.mis complement

<220>
<221> primer_bind
<222> 117411..117429
<223> 12-586-414.mis

<220>
<221> primer_bind
<222> 117431..117449
<223> 12-586-414.mis complement

<220>
<221> misc_binding
<222> 7147..7171
<223> 12-809-119.probe

<220>
<221> misc_binding
<222> 16357..16381
<223> 12-805-115.probe

<220>
<221> misc_binding
<222> 26298..26322
<223> 12-790-396.probe

<220>
<221> misc_binding
<222> 31100..31124
<223> 12-791-211.probe

<220>
<221> misc_binding
<222> 33796..33820
<223> 12-803-125.probe

<220>
<221> misc_binding
<222> 34243..34267
<223> 99-33040-321.probe

<220>
<221> misc_binding
<222> 35281..35305
<223> 12-810-77.probe

<220>
<221> misc_binding
<222> 45880..45904
<223> 12-787-103.probe

<220>
<221> misc_binding
<222> 54485..54509
<223> 12-793-383.probe

<220>
<221> misc_binding
<222> 56570..56594
<223> 12-792-233.probe

<220>
<221> misc_binding
<222> 60324..60348
<223> 99-41009-244.probe

<220>
<221> misc_binding

<222> 60457..60481
<223> 99-41009-111.probe

<220>
<221> misc_binding
<222> 79164..79188
<223> 12-593-174.probe

<220>
<221> misc_binding
<222> 80004..80028
<223> 12-589-152.probe

<220>
<221> misc_binding
<222> 87451..87475
<223> 12-785-200.probe

<220>
<221> misc_binding
<222> 87644..87668
<223> 12-785-393.probe

<220>
<221> misc_binding
<222> 101590..101614
<223> 12-588-103.probe

<220>
<221> misc_binding
<222> 104379..104403
<223> 12-603-191.probe

<220>
<221> misc_binding
<222> 117418..117442
<223> 12-586-414.probe

<220>
<221> misc_feature
<222> 20352,102280
<223> n=a, g, c or t

<400> 1
ttggtagaat tggtagctaa aaggctgagt gaaatatggc ttaaaagctt tattggctgg 60
gcgcggtggc tcacccctgt aatcccagca ctttgggtgg cttaggtggg tggataacct 120
gaggtcagga gtttgagaac agctgaccaa catggtgaaa ccctgtctct actagacata 180
caaaattagc caggtgtggt gatgcatgcc tgtaatccta gatacgtggc aggctgaggc 240
aggagaatta cttgaaccca ggaggcggag gttgcagtga gctgagatcc taccattgca 300
ctccagcctg ggcaacaaga gcaaaactcc gtctcaaaaa aaaaaaagct ttatcattta 360
ttttttggcc ctgtcttatg gtgcagaggc ttaaaagttt tttgacagca aattttctag 420
aggctaggag tgtttattat aaccatgttt ttgagcggtg aggactacct cagagggcat 480
gccttgtgtc atcattgttc ttattgctga gctaccgaaa cctagaatct gactcacaca 540
atatgacact tatttccgtt ttcttggtag agtttgtgtg gtcattcatc ttttagatct 600
tttagaaact acagtcctcc tagttcccac ttttatattt atttatttat ttatttttga 660
gatggagtct tgctctgttg cccaggctgg agtgcagtgg tgcaatctcg gctcactgca 720
acctctgcct cccaggttca agcaattctc ctgcctcagc ctcccaagta gctgggatta 780
caggtgcgtg ccaccatgcc tcgctaactt ttgtattttt tgtagagacg gggtttcacc 840
acgttggcca ggctggtcgc aaactcctga cctcaagtga tcaacccccc tcagcctccc 900
aaagtgctgg gattacaggc gtgagtgacc gtgcccagcc atcagttccc acttttaaat 960
gagaggttct ttgtttttgt ggggtttttg ttgttgttgt tgttgttgtt gttgttgctg 1020
ttgttacttg agacagggtc ttagtctgtc atccaggatg gagtgcagtg gcacaatctt 1080
ggctaactgc agcctggacc tccctgggct caggtgatct tcctacctaa gccccctgag 1140

tagctgggag tgcaggcaca caccaccatg cctagctaac ttttgtattt tttgtggaga 1200
tgagatttca ccatgttgcc caggctggtc tcaaattcct gtactccagc gatcctcctg 1260
cctgagcttc ccaaagtgct gggattataa gtgtgagtca ccgagccagg tcataaagac 1320
agtttttttt tgttttgaga tggagttttg ctcttttgta cccaggctgg agtacagtgg 1380
tgtgatcttg gctcactgca acctccacct ctcgggttca agcaattctc ctgtcttact 1440
tagcctcccg agtagctggg attacaggcg tgccccacca cgctgcactg attttttata 1500
tttttaacag agatggggtt tcaccatgtt ggacaggctg gtctgcaact cctgacctca 1560
ggtgatctgc ccacctcagc ctcccaaagt actgggatta cagatgtgag ccactgtgcc 1620
tggccaataa agagaatttt aaacaatgcc tacagtgtga ttttagtggg gtggccacat 1680
acatgtaggt gtgcacatgt atttccatga aaatgccttc ctgcctttag ccttgacaat 1740
gctgcatata gggaataagc atttaaatga tggtgtctgg gctaagagtg tgaactctgg 1800
agccaaaccc tccaaaattg catcctgaat ctgccactag ctagctctgg agccatgaat 1860
aacttattta acctctctgt gatggtttat ctatctgtaa aatgggggta atgaaacatc 1920
tggctcattg agtgcttggg agcatcaaat aggttagtag gtctggtgtg gtggctcaca 1980
cctgtaatcc cagcactttg ggaagctgag gcatgaggat ggcttgagcc caggagtttg 2040
agaccagcct gggcaacata gtgagactcc catctctaca aaaaagaaaa ataatagccg 2100
gtgtagtggc atgaacctgt ggtcccatct actcagaggc tgaggtggag gatcactgga 2160
gcccaggagg ttgagtctgt agtgaattgt ggttgtacca ctgtactcca gtctgggcct 2220
cagagttaga tcctgtctca ggaaaaaaaa aggttcgtag acagaaagta cttatggctc 2280
agagtaagtg ttctgtgaat gttggctgtt ttttaaatct tgcatttaat tcctctttgg 2340
ttagaagtat acttctttgc ttggattacc atttgcatgt ggagagatgt tccttctaga 2400
accatgtctc cttgtcacat gattatgtac agtgtgagct ttattaatac tctgtcagca 2460
tttgaggttg gaacataaat gcgtggctgt acccagtgga atccagcaag agagagaact 2520
aagcaccgta gcaactgttg ctatttgtgt actgagcaac agctgtgcga gacagattta 2580
ttttttaatc ctgatctcac tgatagcctg ggatgtggcc ttgccacatc tcaaaggagt 2640
agattggttg agatatccag tctttgagga gatttcatga ccacatgcca taaagattct 2700
gtttcatttc aaatatgagg agctaaagct ttgtgtgtat gagtgaggga gaaagattaa 2760
gctttaaagc attcatagtc agtccctgaa atatagtgtg actggattaa tgacatgata 2820
aatcactttt aatgtgggtg gtgagacttc atgcatttca gtcaaaaaag taattttgct 2880
gtaatgtggt tcatttccaa attttggctg ctaacttacc taggatggat aaacattatc 2940
acagagtatt tactgagcat ctactttata ccagggtctt ttctgaatac tggtaacatg 3000
cagttaatta cactccaaaa tgacttagtc tgtgattcca ctatatacta cattacataa 3060
atacacacac actaatcatt tacatacttg tcaaccttag atttcagaaa taacactgta 3120
tttgtaatga atgcttcatt accaaatgct tttctgaagt acatagtaag atctttaaaa 3180
agagctttca tggccaggcg cggtggctca cgcttgtaat cccagcactt tgggaggccg 3240
aggcgggcgg atcacaaggt caggagatca agaccatcct ggctaacacg gtgaaacccc 3300
gtctctacta aaaatacaaa aaaattagct gggcgtggtg gcgggcgcct gtagtcctag 3360
ctactcagga ggctgaggca ggagaatggc gtgaacctgg gaggtggagc ttgcagtgag 3420
ccgagatcgt gccactcccc ttcagctagg gcgacagagg gagactctgt ctcaaaaaaa 3480
aaaaaaaaaa aaaaaaggct gctttcaaat agtggttgga ttcattcaat aactttctaa 3540
ttttcatttc tctttttttt cttgagatga agtcttactc tgtcacccag gctggagtgc 3600
agtgcattaa tatcagctca gtgcaacctc tgcctcccgg gttcaagtga ttctcctgcc 3660
tcaacctccc cagtagctgg gattataggc acacaccact atttctggct aatttttgtg 3720
tttttggtag agatggggtt tcaccatgtt cgccaggcta gtctcgaatt cctgacctca 3780
agtgatacac ctgccttggc ctcccaaagt gctgggatta caagtgggag ccaccgcgcc 3840
cagctccatt tctgatgtgt tccaatattt cagggcttaa cgggtgcctt gaaatagtaa 3900
tgtgggccag gtgcagtggc tcacgcctgt aatcccagca gttagggagg cggagatgcg 3960
cagatcacct gaggtcagga gttcgagacc agcctggcca acatggtgaa accctgtctc 4020
tactaaaaat aaaaaaatta gccagttgta gtggtggatg ccagtaatcc cagctactca 4080
ggaggctgag gcgggagaat ttcttgaacc cgggatgagg tggttgcagt gagccgagat 4140
cacaccactg cactccagct tggggaatag agcaagactc tgtcttcaaa aagaaaagaa 4200
atagaaatgt cgttcatttc caaattttgg ctggtaactt acctaaggag gataaacatt 4260
gtcacagaat atttactgag catctacttt ataccagggt ctgttctgaa tactggtaac 4320
acgcagttaa tgctgttggt ctggtttttc atgttccatg tataaattac ctcttaaaac 4380
tatgattgtt ttcattcatt cattcatttt tgagatcggg tctcaaaata ggctggggtg 4440
cagtggctcg ataacagcta actacatcct caaacctcca ggctcaagcc atcctccgac 4500
ctcagcttcc taagtagctg ggactgcagg catgtgccac catgcccagc tagttgttgt 4560
tgttttttt ttttttggta tttgttgtag agacagggtt tcaccatgtt gcccagggtg 4620
gtcttgaact cctgacctct agtgatccac ccagctcgga ctctcaaaat gctgaaatta 4680
caggcatgag ccactgagcc cagcctatta atttttgatt gccctattca gatttagtgg 4740
tatgaattta ctgattggga tgttttgaac acacagaata acatgtttaa ctaaagaact 4800
tgtaattgag tgacttataa aatgaaacat ttttatcttc taggtattat taacccaaag 4860
aatccaaagg aagctccaaa gtccttcagc ttcgactatt cctactggtc tcatacctca 4920

gtgagtaccc tcatgccaca gcactgccag ctcctgcctc ctttcctctt tccttgccga 4980
tttgtctttt ctcttagctg agcagatcca agttttgctt cttaactgtt gtgtagcatg 5040
agttcattct ttgatacagt gatatagtat tttcattcat aaagtatttg ggttgatgga 5100
aagtttatgt acacatatac ataaatatat acacatatat acatacagaa catatttagt 5160
ttacaccaaa tttatactga atgtacacta aatttatact gatttagaca tttcttgagg 5220
atttatttat cagtatctta ggatcctcat taccaacatt attgcccatt gatcttagga 5280
aatgccaaat aataataatt attattatta ttattatttt cttctgagac ggagtctcac 5340
tctgtcgccc aggctggagt acagtggtgc catcttggct cactgcaacc tccgctcgag 5400
gttcaagcga ttctcccgcc tcacctcccg agtagctggg attacaggcg cgtggcacta 5460
tgcccagcta atttttgtat ttttagtaag acaggatttg accatgtggc cagctggtct 5520
cgagcttctg accttagtga tccacctgtg tttgcctccc aaagtgctgg gattacaggt 5580
gtgagccact gtgcctggct ggaaatgcca aattaaaatt gcttagggta ctgttaattt 5640
taagttgaat tttgagggaa tggccagcaa gattggaatg accacagaga aattgttttt 5700
tttttgtttg tttgtttgtt tttggctagg ctgatcttga actcctgacc tcctgatcca 5760
cctgcctcga cctctcaaag tgttgggatt aaaggagtga gccaccgtgc ctggccaatt 5820
ttttttttt aattgtagta gagcaatgtg ggttctgaac attttgaaag atgataccgt 5880
gcctctacta aaaattccaa aaaaaaaaaa aattaggtgg gcttggtggc gggtgcctgt 5940
aatcccagct acttgggagg ctgagggtca tgacctcgcc gtaaccgcat ggcgtcccac 6000
actgcactcc agcctgggtg atggaacatg actgtgtctc aaaaaaaaaa aaaggaaaga 6060
ggctgcagat ccacaagaaa ctaaaagatt accagtatta cgaaaatgca gagtagggcc 6120
cttcctggta cccttactca gtgtggaacc ccaaatgttg ctgcttctta gggattgttc 6180
atagagttta cttctgcatc attcatctct ttgttacatg tttttggttt tgtttttcct 6240
gttttgtatt tttttttaaa gactgagtct cgctctgttg cctaggctgg agtgcagtgg 6300
catgatctca gctcactgca acctacacct cctgggttca agcagttctc ctgcttcagc 6360
ctcctgagta tctgagatta caggcaggtg ccaccacacc cggctgattt ttttaatttt 6420
tagtagagat ggggtttcgc tgtgttggcc aggctggtct cgaattccta tcctcaagtg 6480
atctgcccgc cttggcctcc cagagtgatg ggattacagg catgagccac tgtgccctgc 6540
cctggggaat cgatgtttta aactattgac aaagcctaat tgtcaaggga agggactgag 6600
aaggggattg gaaatgctgt ttcctgaaag ctgtgtttat tgagaaatag tcccagtagc 6660
tggagtgagt tcaaagaagt atttttatat gtggacttga cctttggtct tttattctca 6720
ttttccactt aagaaaatct tggctgtgtg aatgaaagag tgatactttt aaggttatag 6780
aaagtgaaat gtaatcatgc cagataattt tatatagata tttttatgta tggctgacct 6840
ggatgactct aacagtgcat gtgtttgtga gtgtgtgtgt gtgcatgtgc gtgtatttaa 6900
tgagaaaagt aaacttgtgt ataggaggct taaaaaatgt gtagggaatt ttaggtgact 6960
gttctgattc cagacacttt tattatggaa gcaatcaagt aagtatagga agaaatatta 7020
ataaaaggtt atttatttct ccttttactc tttacagccc gaagatccct gttttgcatc 7080
tcaaaaccgt gtgtacaatg acattggcaa ggaaatgctc ttacacgcct ttgagggata 7140
taatgtctgt atttttgcst atgggcagac tggtgctgga aaatcttata caatgatggg 7200
taaacaagaa gaaagccagg ctggcatcat tccacaggtg aaaaacaaaa caaaacaaaa 7260
atcttctctt cattattagt gttagtctta aattgcttta acagttattt ttatttggcg 7320
aacatttatg cggggattgt tttatgtcag gcacaaagat gaacaaccca ttatttccc 7380
tcagaggagc tcacaattga atgggaagga ttgacatgta cacatgtctg tcattaaagg 7440
tgggaaagtc agtgttttgt aatgattttg ccatatatcc aatgccatat tattttgtca 7500
tttgaaaagt gttaccagct tgttaaagct ctgttctaag tcctggagat gggggggatat 7560
tgttgatctg attttttttc aaattccatg catagattac ctctgaagaa tgtgattatt 7620
tttggttttt ttgatagcct tattcggaga tactttcatt ttatttttcct aatttaatac 7680
tatgacttta aacttcaaac actctgagat tctccctctt ttttttttt tttagaacag 7740
ttagtattat tgggtttgat cctttattgt ttgggaatga agaagcgttc tgacatagat 7800
atttatttat ttttattttc attattcgtt tcacttagga tttaatggtg aaagacataa 7860
gtatttaaaa agaacttcat tatttactta tttatttgag acagggtctc actctgttgc 7920
ccatgctgga gtgcagtggt gtgatcttgg ctcactgagg ccttgtcctc ccaggctcaa 7980
gtgatcctcc tgccttagcc tccagagtag ctgggactaa agccatgtgc caccacacct 8040
agctcatttt taaattaaaa aaaaaaaaaa ttttttttt tttttgaggt ggagtctctc 8100
tctgttgccc aggctggagt gcaatggcgc catcttggct cactgcaacc tccgcctcct 8160
gggttcaagt gattctcctg cctcagcctc ccaagtagct gggacaatag gcacccacca 8220
ctatgcctgg ctaatttttg tatttttagt agaggcgggg tttcaccatg ttggtcctgg 8280
tctggatctc ctgacctcag gcaatccgcc cacctccgcc tcccaaagtg cagggattgc 8340
aggcatgagt caccatgccc ggcctttaaa attttttttt tagagacagg gtctcaccat 8400
gttgtccagg ctggtctcga atggacctgg gctccatcct gggcttggga tgatctgcct 8460
gcctcgactt ccagaagtac tgggattaca ggcatgagcc actgtgcggg gcctaaaaac 8520
ttaagaaaaa aaaaagatta agtgagaaga ttgtatttaa aattttcctt ttatagtcat 8580
gcatcactta agggtggggg tacattttga gaaatgcatc attagtcgtt aggtgctttt 8640
gtcattgtac agatatcata gagtgtactt acacaaaccc tggaaggtgt aaactgctcc 8700

acacctagac tgcgttgtat agcctgttgc tccaaggcta taaacttgtg tagcatgcta 8760
ctgtactgaa tactacaggc aattataata ggatggtaag tatttgtata tctaaacata 8820
gaaaagtaca ttaaaaatat gatataaaag atttgaaatg gcacacctgc atagtgctct 8880
tacactggag cttgcagggc tggaagttgc tctgggtgag tcagtgagca agtggtgagt 8940
aaatgtgaag gcctaggaca ttgctttgca gtattgtaga ctttataaat gctgcacact 9000
taggctacac taaatttagt tttaaaaaca gtttcctttc aataattaac cttagttaaa 9060
ttactataac ttttttactt tataaacttt aaaatttttt aaactttctg atcctttttat 9120
aataaccttt agtttaaaac acaaacagcc ccagcatggt ggctcatgcc tgtaatccca 9180
tcactttggg aggccaaggc gggtggatca ctggaggtca ggagttcgag accagcctga 9240
ccaacatggt gaaaccccgt gtctactaaa aatacaaaaa aatgagccag gcatggtggc 9300
gcatgcctgt aatcccagct acttgggagg ctgaggcagg agaattgctt gaacccagga 9360
gacggaggtt gcagtgagct gagatcgcgc cattgcactc tagcatgagc aacaagagtg 9420
aactcggtct caaaaaaaaa aaaaagaaac cacacacaca ctttgtacag ctgtacaaaa 9480
atattttctt tcttcatatc tttattctat aagctttttt gtatttttta attttttttt 9540
tctttcttga gacagagtct tgctttgtca cccaggctgg agtacagtgg cgccctatca 9600
gctcactgca acctccgcct cctgggttca attaattctt gtgccttagc ctcctgagta 9660
gctgggacta caggcacgta ccaccacgtc tggctaattt ttgtattttt agtagagacg 9720
gggtttcaca atgttggcca ggctggtctt gaactcctgg cctcaagtga tctgcctgcc 9780
ttggcctccc aaagtgctga gatttacagg cgtgagccac tgcacctggc caaatttttt 9840
ttttttttt ttagacagag tttcactcct cttgcccagg ctggaatgca gtggcacgat 9900
cttggctcac tgcaacctcc gcctcctggg ttcaagcgat tctcctgcct cagcctccag 9960
aatagctggg actatttttg tatttttagt agagatgggg tttcaccatg ttggccaaga 10020
tgttctcgat ctcctgacct tgtgatctgc ctgcctcgtc ctcccaaagt gctgggatta 10080
caggcatgag ccaccttgcc cagcccaaat ttttaatttt tgaaaaaaaa tttaaacttt 10140
tttgttaaaa acttaggcac acatttatta gcctaggcct acacagggtc aggatcatca 10200
agatatcact aggtgacagg aattttccag ctttcttata atcttatggg gatcactatc 10260
atatatgcag tccatcgttg accgaatcat ggttatgcag cgcatgactg actttattca 10320
aagcattaat cgtatcttga tgtttatgac ataatatatt ttgagatgga gcaagaaagg 10380
acgtctggct aattcattga gcacgcgtgg atgacttagt actcctctca tttgtgctct 10440
tcatgcctct ctcattctac ttccctagtt atgtgaagaa ctttttgaga aaatcaatga 10500
caactgtaat gaagaaatgt cttactctgt agaggtgagt acagccgtga gttgacaccg 10560
taagcccttg ttttccattc tctcaagcat cacttaaatg gctccaaatt atgactgtgg 10620
tacacatcac ttcaccattc cttcattttt gtcctttcag gctatttctg ggttttgggt 10680
ataatgaata gtgtcctgct ttaatctttt ttttgggggg tgcgcggtgg ggacagggtc 10740
tcactctgtt gcccaggccc tggagtgcag tggcgtgatc acagctcact gcagccttga 10800
ccttctgggc caaagcgatc ttacccccctc agcttctcaa gtagctggga caacaggtgt 10860
gtgccaccac cacccctaga taattgattt ttttgtagag atggagtttc cctatgttgc 10920
ccaggttggt gttgaactct tgggctcaag caatcttgcc tcctccacct ctgaaagtgc 10980
tggggttgca gtgtgaatca ctgagcccag ccacaattaa atctttgttc ataagtctgt 11040
gcatttattt gcagttatta tcttttgtat gtttcttggt ttttaaataa tacatttgaa 11100
ttgccaaatg ccttttgga aagtaatgct gggtaacagt tttcagtcac actaggaggg 11160
tgtaagcacc cattgtgccc ttgccaacat ttttctattt tttaatcttt gctaatttga 11220
taatgtaagt attctcttac ttaaattttg aatttctttg attactagta tggttgaacg 11280
tcttgtaaat tatgcatatg cctgtttaca ttttgggttc acaatgtctt tctgattcat 11340
ttgtgggaac tcttctcctt taatattgtt aataatggtt tttgacaggc attggtttta 11400
aatatatata atcaaatctc ccccttccc actttttaa ttgcttataa aaacttggaa 11460
agtaattatc catcctgtgt ttaccctggc tctttcatga agtctataga attactttat 11520
acacatttgc ttatttgata aacattctct ctgttttgta tgtgcttaaa gaagctgttt 11580
tggggctcat acctgtaatc ccagcacttc gggaggcgaa tggatcactt gaggccagga 11640
gtttgagacc agcgtggcca acctggcaaa accctgtctc tactaaaaac acaaaaaatt 11700
agccggctgt ggtggtgcac acctgtaatc ccagctactc gggaggctga ggcacgagaa 11760
tcgcttgaac ccaggaggca gaggttacag tgagccgaga tcatgccagt gcaccccagc 11820
ctggagtctg tctcaaaaaa aaaataataa taataaataa ataaaaagaa gaagctgctg 11880
tttctttctc ttcctccctc tctctctcac cgccaccccc ccaccacacc ctttttggt 11940
attcactttc atagtaactg aaattattaa agtgttctgt taccatgagt ctgttttgaa 12000
aaaataggct gccttttaag cgatcttgag attttttttc ttttctttt ggtatttgtt 12060
ctgattctct gtagccttac agcatttcat tcctagttaa ttcatttaat tgctttggag 12120
ttttttcttc acctagaagg tgttgcataa atatgtaaat tctaaggaat tctgtcttgt 12180
gactgtcata caccaaaaag agaatgtgct gccctaagca tattttgcgt taaaatgaat 12240
tttctgaata gcttgtccca tagttattct tttttgaatc aggtgataga atttttgagc 12300
ttacaggtaa taaaagaaga tgccctttc ttttgtatta gagtttccag gcttaacttg 12360
gagtggagca tagattccta gttgacactg tggtttcttg ccacatttgc aggacctctc 12420
tacctgtctc ttctacttcc ttccctgttt actctgctcc acccatattg cctttttttcc 12480

tgtttcaaga acataccaga aacattcctg cttcaaggcc tctacattta cttctcctct 12540
tgcctgaaat gttgttcact tggatgtcat tttgactaag tttctcacgt tcttcaagtc 12600
tgcataaatc acacactctc agtgaggctt accttgacca ccttatttaa tattgtatct 12660
tgccctttac cccacatcct gagctctctt ccccattcta ttttcttttt tccatcatac 12720
tcatcatgtt ctaacatgct ttctcagtta cctatttttt atgtttgttg tttattatca 12780
gtatcccttg ctagaagcat aagctcactg gggcagggtt ctttgtctgc tttatttagt 12840
ggtgtatacc aattgcctag aacagtgcct gtaagagaac ggtcctcagt gagttggatc 12900
tgccaggtgg catctggagt ggttggtgca gaagtaaaag aaatgatgat ggctttggat 12960
ggattcacat atcagagcat aaggaatgga gcacatgaac ttttacctag attatttggg 13020
ttctctttgt ctcacatttg tgtgaaaagg tacgagtgtt ccaaaagatg taatatattc 13080
cggggattca gtattgtatg tacaaattta agtacacagc aatgattgaa gaatcatctc 13140
actttctagg atgacttgct gtcagaaaga atatttgttg ttctgttact ttactctcat 13200
aatagattaa ttactttatc tttcagagtt gtctttttt ttggagacag agtctcactc 13260
tgtcacccag gctggagtgc agcggcgcgg tctcggctca cttcaacctc cgcctcccag 13320
gttcaagcaa ttctcctgcc tcagcctccc gagaggctgg gattacaggc tcctgctacc 13380
acacctggct aatttttctgt atttttagta gacacagggt ttcaccatgt tggccaggct 13440
ggtctcgaac tcctgacctc atgatccacc cgccttggcc tcccaaaggg ctgggattat 13500
aggcgtgagc cactgcgccc ggccttttt tctttttctt ctttttttt aagacaaggt 13560
cttgctctgt tacccagact ggagtgcagt ggtgcaaaca gggttcgctg cagccttgac 13620
ctcctgggct cagtgatcgt cctgcctcag cctcctgagt aatttgcacc acaggcatgt 13680
gccacagtgt gtggctccca aagtgctaag attacaggca tgagctactg catctggcca 13740
agttcttaat tttttaaaat ggattttatt tatttattta tttatttatt tatttattta 13800
tttatttatt tttagagaca ggatctcact gtcacccagg ctggagtgca gtggcgtgat 13860
ctcagctcac tgcatcccca acctcctggg cttgagtgat cttcccacct cagcctcccg 13920
agtagctggg aacacaggca cacactacca tgcctggcta attctttgta tttttggtag 13980
aagtggggtt ttaccatgtt acccaggttg gtctcgaatt cctgagctca agcaatctgc 14040
ccaccttgac ctcccaaaga gtggggattc caggcatgag ccaccatgcc cagcctttaa 14100
ctgattttaa aacacaaatt aaggccaggt gcggtggctc acacctgtaa tccctgcact 14160
ttgggaggcc aagatgggag gattgcttga gcccaggagt tcaagaccag cctgggcaac 14220
gtggtgagac cctgtctctg caaaaaaaaa ttttttttaa ttagctaggc atggtggcac 14280
gtgcctgtag tcccagttac tccagaggct gaggcaggag gattgcttaa gcccaggaag 14340
tcgaggctgc agtgagccat gtttgagcca ctgcactcca gcctgggtga cagagtaaga 14400
tacagtctca aaaaatatat aaaaataaaa tacagtataa aaatacagtt agagactgaa 14460
tgaaagaaaa aggcaataga aatggaatta agattccatt ccagatactc tgaaaacatg 14520
tccatagaca tttaaagttt aaaaatgcag gagctgtaat gtaagcttga atagctccat 14580
gatacagaag gatctgtttc ctgggtagca tcgtaatgtg gcctattgtg atgaatttac 14640
cctagttcag agccttgttt gcttagaagt ctgattttgt agattcatca agaaaaggac 14700
tatccttaca agttaagact tgagtaaagc tcatttgact gatgttatga aaccaacttg 14760
ttggtttgtt actttaccat gagaaataac tattactaag gctagaacac tcaagccatt 14820
cgcaagacta tttaatatat ttctgtggac agacctcgct gttggaagcc tgttggttgt 14880
ttggtgtgca tggaattaga aggagaggag ttgcttagat acatgtttaa ttgccccagg 14940
ttaagaaatg atagcttaat gcattaatca ccatgatcga atggtcctgc caatgagaca 15000
agaatttatg aagtgtattg tctgaggctt ctttctccct cacaaatctt ggataggtcc 15060
tgccaccacc agctttcatg ctggtcaacc ctcatccctt ctggaggata catcgtccat 15120
agttagtaac agtgtggtga gccttaaaga attagcatca cccgtctgtt ttgtcttctc 15180
tgctctgtaa tattgatatt gttccttggt tttaactcta atagaaatct tatttccgag 15240
ttcttggagt aacctcaatg caaatcccag ttactgacaa cagtatacat tttaaaaaac 15300
attgataact gataaattgc tactaaaata tcttttagct gtcactacat tccagagata 15360
atctgtttac ctctccctta cagagaaagt ctgaaggctg gttctaacct attcagtgta 15420
attacctgtg cgtattgtta atagaataaa agggcctaga gttacggcat ttcactttgt 15480
aaaggttgct gatctctaat cttagagctt ttaataaaag ttaaaaaaat tttttttctg 15540
tatgtgattt cttttgtttt ctttttcttt ctttctttct ttctttttt ttattaagac 15600
agagcctcac tctgtcaccg aggctgaagt gcaggggtgt gatatcggct cactctaacc 15660
tccacctccc aggtacaggc gatttctcct gccttagcct cctgagtagc tgagactaca 15720
ggcatgtgcc accacgcctg gctccttttt gtatttttgg tagagacggg atttcaccat 15780
gttggccaga ctggtctcga actcgtagtt cgagtgatct caagtgatcc gccctccttg 15840
gcctcccaaa gtgctgggat tacaggcgtg agccaccgcg cccggcttct gtatgtgatt 15900
tctttttcac tctaattcac tttactaatt tgttcatagg tgagctacat ggaaatttac 15960
tgtgaaagag tacgagattt gctgaatcca aaaaacaagg gtaatttgcg tgtgcgtgaa 16020
cacccacttc ttggacccta tgtggaggat ctgtccaagt tggcagttac ttcctacaca 16080
gacattgctg acctcatgga tgctgggaac aaagccaggt atggtaggaa atagagtaat 16140
gactgaggtc tttggcacct tttgaggtcc ttttttccca gttaagggtt tgaggccaca 16200
tttatagcta tgaaagttgc tttaattgtg gagtcctctg atcctttgac tgtgctgtaa 16260

cagtgagggc ttcagagtgt taagtatctt ctcatctaaa cagtacattg ctgaaacatg 16320
ttagagctct cattgagttt tactcttagg ttttgatcaa gtcatggtra aatgttagtt 16380
atcatgtata tatgcggctg tggtagtaat tgattctgga ataaatgagc accaataaag 16440
aataaaaatt aaatgcttca ttggtttcta aagggatagc tggagtatga aaaatgattc 16500
tgatggggtg cctttggaat tggagagtaa cttgatgggc tcacaaataa ggcagtattg 16560
agagcaattg tgtaagtgac aatcttgcaa tctgtgaata agtttaagac tattgcttgt 16620
gattgttatg cttataatgt ggagcctgct gtccatttca actctcatct aagaattccc 16680
ttgtcacgtg gtcaccttta tgtttattta ggacagtggc agctacaaac atgaatgaaa 16740
caagtagccg ttcccacgct gtgtttacga ttgttttcac ccagaagaaa cacgataatg 16800
agaccaacct ttccactgag aaggtaggag agtttcagtc tctaggcttg agttgtgaag 16860
gatggagatt ttgagaagtc ccttttgttg ccctctgctt tttgtggaag gttgggtgtt 16920
ggggggtgct cttttactta atggagggtg ttttatacct ctgcttatca tttttcttcc 16980
accctacacc agttttctgt ataggttctt aaatgaagtt gtaagataga gaatagggtt 17040
aagaaaaaaa caatccagga ctttgggtct caaaaactac tcaatacaag acagttaagc 17100
agtcttgcag gagaaagttc tttcttttct tcttcttgac tgaacctttt tattacctta 17160
ttagctgcct ttggtatgtg tattcttttt ttttttttt ttgagctatg ataatgttat 17220
agtgggatta aatgtttaag caaatgatat gtttcattcc tggaaaataa ctccacaaca 17280
gttctggtgt tcttcaaata cctttctttg cttattgatt tttcattttc tatagtggtt 17340
aaatgtcaca gactttagcg ttggattgct tgctttgaat tctgattctg ccactcttca 17400
gctatattac tttgggcaaa ctacttaacc tctttgttcc tcagtttccc catccataac 17460
agcctgatag gattgttaaa aagattaagc gttaatatag gttaagtgca tacaataatg 17520
tctggcacag agtagccact caataatgtt aattttattg ttattgttag taataaaaaa 17580
tttatgttga atactactct atctgcatcc ctaagaagga tgcacctgga tttttaagct 17640
tcaagttagg tttattgtac tggcaattat aaaagatcct ttgccttata aaatgagtac 17700
tttagacacc actgtcacta attattaaaa tcacctttta tttgttgact tatttctagc 17760
tgaaaatgat ttttaaaaat atactcctgc tgggccgggt atggtggctc acacctgtaa 17820
tcccagcact gtgggaggcc aaggcagaca gatcaagagg tcaggagttt gggaccagcc 17880
tggccagcat ggtgaaaccc agtctctact aaaaatataa aaaaattggc tgtgcgtggt 17940
ggcatgccct gtggtcccac tgctcagagg ctaggcagga gaatcgcttg aacccagagg 18000
tggaggttca gtgagccaag atcgtgccat tgcactccag catgggtgac agagcaactc 18060
tgtctcaaaa aaaaaaaaaa acacactcct actgggcatg gtggctcaca cctgtaatcc 18120
aagcactttg ggaggctgag gcgggtggat cacctgaggt caggagttcg agaccagcct 18180
ggtcaacatg gtgaaaccca gtctctacta aaaatacaaa aaattagccg gtcatggtgt 18240
gtgccggtaa tcctagctac tccagaggct gaggcaggag aatcgcttga acctgggagg 18300
cagaggttgc agttggctga ggttgcgcca ttgcacttca gcctgggtaa caagagcgaa 18360
acttcgtctc aaaaaatata tatatatttc ttatatagat gatcaaataa cttgtttagc 18420
cagaggttat aatattagtg ctatttatta gttagtgttt atctgaatgc cagcaaagaa 18480
cattgtatgt gaaggtgttt tttgcccttc tcccatgatg tcgaatctac aaaattaatg 18540
ttaatcccat aatcatagct ttcctggtaa ttcattttgg atacttttgt ctaaactttc 18600
ctgaagcttt attttttttac agtgttattg aggtattatt gaagtataca gtaaatggca 18660
catatttaaa gcgtaaaatt tggtcagttt tgacgtattt gtatacctgt gaaactagcg 18720
ccacaatcca gataatgata atttttattc tcaaaagttg cctcatatct ctttgtagtc 18780
tttctctatc ccttcttctg tggcaattcc cgggcaattc cggacctgcc cttttgtcac 18840
tatatattaa tttacatttt ctagaatttt atataaatgg aatcatagag tatgtactgt 18900
tttttgtctg gagtctttca cttagcatag tgattttgag attcatctat gttgttgtgt 18960
ttatcaatag ttccttttca ttgttgatga gtattgcgtt atacagatac accacagttt 19020
attaattcaa ctgttgatgg acttttaaac tgttgccagt ttttgattat tacaaacgaa 19080
gctgatgtga acatttgtgt acaagtcttt gtgtggacat ttgtttttgt tctgttggta 19140
aatacttagg agtaggatga ttgagtctta ttgtagctcc ttaagaaatg tcaaggctgg 19200
gcacggtggc tcacgcctgt actcccagca ccatgggagg ctgaggcggg tggatcacaa 19260
ggtcaggagt ttgagaccag cctggtcaac atggtgaaac cccgtctcta ctaaaaatac 19320
aaaaaaaatt agctgggcgt ggtggtgggc gcctgtaagc tcagctactc gggaggctga 19380
ggtgggagaa tcgtttgaac ccaggaggtg gaggttgcag tgagccaaga ttgcaccatt 19440
gcattccagc ctgggtgaca gagcgagact ctgtctcaaa aaaaaaaaaa aaaggaaaga 19500
aaataaatgt caaattgttt tccaaagtga ttgtaccatt ttacattccc accagaaata 19560
tgtgattgtt ccagttgctt aaagcttcac ttttagtaag ccatttaatg agttgatgag 19620
ttttatatat ttttataaca tactctagta aattatgtca tgtttatcct aaagttactt 19680
ttttttgaac cgtgataatt gctatgaatt ttgtgcatgc tctttttttt tttaattaaa 19740
aaaaattttt ttttagagac acaagggtgt tatatgttgc caggctggag tgcagcagac 19800
agctttgaac tcctgggctc aaaggagctt cctgagtagc tgggactaca ggcatgtgcc 19860
attgcacctg ccgtgtgtac tgtttaaact aactgataga tgttgtcttg atctttttgt 19920
gccatattta aaagctctaa tacttttagt tttttccata ccttaatctt ttaaaaagca 19980
ttctgcctct aaatatttcc tgcttctatc ttgcttattt ttgaattgcc cccatgttat 20040

tgttctttat caggtcagta aaatcagctt ggtggatcta gcaggaagtg aacgagctga 20100
ttcaactggt gccaaaggga ctcgattaaa ggtatttatt ttagcaaata aatggcctga 20160
tcaataaatg gccactacct gtttttgtaa ataaaatttt attgaaacac agctacatac 20220
atttgtttat gtattgtcta tggctgcttt tgcactacaa tggcagtgtt gtgacagaga 20280
ccgtatggcc cagaaaacct aaaataattt gtgtctgtcc ttctacggaa aaattttgct 20340
gactcctgat cngtaacttc aatggaagga taagttatct aactaaatag ttagggagca 20400
catcatagtc agggtgagag ttgaattgca ttttaaggaa caggtagaat ttggctagat 20460
ggaattaaga gatggcagtg ggaataggat atacaagatt cagatgtgag aacagaaatt 20520
tcatgtcgat aagttgaggg aagaaggctg cacacatctg taagcttttg gagcacaaat 20580
atctcgaaga tacttcagga ctgtgagtta ggcatgtggt ctcagcatgt tattatcaag 20640
tggaaaagag gacagatgga ggactaaagg ttgtttttag aacttgtaat aaagaaatat 20700
gctagcatat ggcaaatcat attttatgtt ctgtatgata tagcagtagt agaaattctt 20760
catcattctt tcatatttgg tatttatttt aggaaggagc aaatattaat aagtctctta 20820
caactttggg caaagtcatt tcagccttgg ccgaggtggt aagttttata cttcattata 20880
aggggagttg tgtctgttca gcaaatatac ttgatgatat tttattattg tttataggct 20940
gtctttttgt gcttttctgt aatctgcctt tctcctttta gagctttca gataagatac 21000
taggtataac acagtttgac tttgtactga aagtaaaaga tactaaggag aacaaaagta 21060
aatatagatc tgtaaaaact agagcaaaca ctgaaggttt gactcatttt tttgagcctt 21120
attttatgtt tctcacaaaa accaagtata tttattgtct cttcttttcc tgtataatct 21180
tttctgagta agatgcagta tcttttggag tattttgagc agtttggctt tggtgaatgc 21240
tgttgatttg agggccagta gttttttttt ttttttcccc ccaatagttt tgttctttcc 21300
tactccttta tggtagaaca gtaaaactaa aactattttt tttaaattag gggctttgtt 21360
gctacattat ttaaattagt cattgtgtgt tctagtttag tttaaaacct ctttttaatt 21420
gctaaaaaaa tgaaatgctt tcttgcttgc cttggagaaa tcataatcta tgacatgaaa 21480
tttaaacaaa atctaatttt ctacttggag gcttatcctg tgttcttatt ttctccttta 21540
aatgctttca cctgtaggat aactgcacta gcaaggtaca gtggggattg gtagagataa 21600
actagaattg acttttatgt tttaaatcct cactaggatg tatggaggca taagtaggaa 21660
tggaaccttc aaaaatcttt tcatcatttg ttctctggct ctggaattta gagtggctgt 21720
aaatttaggg tgaccacaaa tcatgaggtt tcacaccaaa attaatttct agaacatgat 21780
atgctttgga gaaggaatta agaataaagg aacaaggccg agcatggtgg ctcatgcctg 21840
taatcccagc actttgggag gccaaggcgg gcggatcatt tgaggtcaga agtttgagac 21900
cagtctggcc aacatggtga aactctgtct ctactaaaaa tacaaaaatt agctgggtgt 21960
ggtagcgcgt gcccatagtg tcccaactat tcacgggttt taaactgaag ttttaggcct 22020
caaggattac atagtggagt gaaagtttgt tccactaatc tgaatgactg aaaggcttaa 22080
ttctcccttg aggtctgtgc aatctaaaat tgaacctgct attggaaatg actttctgat 22140
tggtggattg tgtgtctgtg ggtagagtgg aatgagatga tgatgatgtg gacatattta 22200
ataagagatg gaattaaaat gtttcttgcc tactcccttt cctcctcctc ctcaaaaaaa 22260
aaaaaaaaaa aaaaaaaaaa aacccaacaa acaccacaag actagccact aagaggccag 22320
ccatgtagta ccctttttct gaatagaaaa agtgcttgtg caatagtagg atcatcctta 22380
gtttaacctt ggattatacc aggactgaat tataaaactg ctgatttgct ggagtggtgt 22440
ttagcagtga gaccctggct ttctccttag tagcttcttt ttttttttt ttttttggag 22500
accaggtctc tcactgctgc ccaggctggg gtacagtggc acgatctcgg ctcagtgcaa 22560
cctccacctt ctgggttcaa gcgattgtcc tgcctcagcc tcccaagtag ctgagattac 22620
aggtgcccgc caccaccccc ggctaatttt ttgtattttt agtagagatg gggtttcatc 22680
acgttggcca ggctggtctc gaactcctga cctcaagtga tccacccacc tcggcctccc 22740
aaagtactgg gattacaggc atgagccact gtgcccggcc tccccttaat atcttcttaa 22800
agtgtcacaa cgcagtagtc agccagcata tatttgagta gctgttacct gtaaagcatt 22860
tgcattaatt tttacttta gagaatataa ttttttttca tttaagcact gccttttcat 22920
tggtgctttt tgctttgtcc tgtggtaata agaaggatca tgaagaaagc agaataagta 22980
gacttaatag aacattcatc tttctaaatt accaaatcat atctctttct tagacctttt 23040
tacttttagt ctttgcttta ataattggat tatttttggc ataatttcct ttcaaagaga 23100
ccaaataaat attattttac aagagttact agatttgcaa caacccaacc aatgacatgg 23160
tggttggcca agtaaattca ctaactagat gacttgctcc tacacatcca acctggaatg 23220
cacaagatta gtattttccc ctgtgcaaaa ataaaaattg aggatggggt atttgatgat 23280
gttagggaat tattgtttat ttaatgcaca agattagtat tttcccctgt gcaaaaataa 23340
aaattgagga tggggtattt gatgatgtta gggaattatt gtttattttg gtaagtgtga 23400
taatggtcct gtggtgtttt gttttgtttt taaaaaacat gacctttcc ttcacagtaa 23460
gaaaaaggat aaaaaatgat atggacagga tattaggtaa tagtattgtg ttgatgttta 23520
atttttgaa tttggtcatt gttttctgtg taagagaatg ttcttatatg ctgaagtgtt 23580
taggggtaaa tggagatgag gtccgcaact gacttcaaaa tgggtcagga aaataagttg 23640
tagaaaagtg tttgggggat aattgaaaga tgattggcag agtattttgg taatggttga 23700
agctgcacgt tgggtatggt agctcattat actgctctct caactgttgt gtatattaca 23760
cagttttcaa aataaaatgt aaaatgaaaa aattacctta ttgcgtgagt ttttggtttt 23820

gtgaagaatt gaaactatga atgataaact tccacagtgt cggggttgta ctgtaattta 23880
atttacttgg gaattttaga ctgatttgcc tttcttggga atttttttcc ctaacgaaaa 23940
atgctaagac catttctttt cctttaatag agtaaaaaga agaagaaaac agattttatt 24000
ccctacaggg attctgtact tacttggctc cttcgagaaa atttaggtat gttgaccact 24060
agtgaaaagt agttatcttt ttaactttga ttatcttttg tggttaattg tcctgtgtct 24120
gttttggaga caaataatct ctagatattc atgttactac ttgagtgttc cttgcctgct 24180
ttcattttta ctcctaattt cagttggttg gattcccttg gttttaagcc agggtatact 24240
cccatatctc attttcttca gggcatgaac tgtgcagtag ttgtcctccc gactccccat 24300
acttgcttca gattattgct ctttcttcga agttatttca aaactccatg cttctctgag 24360
gtgcacaaaa ttagacattg ccaaccctta atcctcctcg ttgctgatgt tttcagacct 24420
tcaaatcact gttcttcatt cacttctcct acttttttt ttttttgag atggagttcg 24480
tgctgttgcc caggctggag tgcggtggtg caatcttggc tcactgcagc ctccgcctcc 24540
tgggttcaag cagttctcct gcctcagcct cccgagtagc tgggattaca gacacacacc 24600
accttgccca ggtaattttt ttgtactttt agtagagaca gggttttgcc atgttggcca 24660
ggctggtctc gaactcctga cctctggtga tccgcccacc tgggtgaaag agtgagactc 24720
cctctcaaaa aaaaaataaa agaaagaaat ttgacattac attaaatctt aggatctgag 24780
attacaatat cagatttgac acattccata aaatttttct tctaaaatga gtgtgatttg 24840
atactcatga ttaatctttt taggtggcaa ttctcggact gcaatggttg ctgctctgag 24900
ccccgcggat atcaactacg atgagacttt gagcactctg aggtactttc ttttgatctc 24960
agtaacaaca tagaccacat tgccatcaga agccattttg tgataccatg gatgttttta 25020
tgctatctgg gtagttattt atgtaaataa tgttcctttc ctctcataac agggaaaact 25080
tagggaaaaa gtgatttgta ttataataga atttacctct ttatggtttt cagtttcact 25140
gggaaagata tgaacttttt attttttttat ccctcatgga tacctagata aggcaggagc 25200
tttttcatag gaaagatttt taaattgctc aagtgagatt tttgggtaac ttttaggtaa 25260
tagtgacatt ttcttattca tttctataaa aagaagtgtc ccccttcttg tagtgtttgt 25320
cgttatagaa atgctgaagt cggccaggcg cagtggcccc cacctgtaat cccagcactt 25380
tgggtggccg aggtgggcgg atcacttgag gtcaggagtt tgagaccagc ctggccaaca 25440
tgatgaaacc ccatctctac taaaaatata aaaatgagtc gggcgtggtg ccatgcacct 25500
gtaatcctag ctactcagga ggctgaggca ggagaattgc ttgaacctgg gaggcggagg 25560
ttgcagtgag ccgagatggc gccactgcac tccagcctgg agacagagcg agactccatc 25620
tcaaaaaaaa aaaaaaataa attctgaagt cataataagt aataaaatta taggaatatg 25680
tagcaatatt aaaaaattac caacatagcc aggtgcagtg gtgtgcacct gtagtcctag 25740
ctacttggaa agctgaggca ggaggatcac ttgagcacag gagtttgagg ccagcccagg 25800
caacgtagtg agactctgtc tcttaaaaaa aaaaaaagaa aaaaacaatt accaacagaa 25860
tgttcataaa aagtttttt tgtttgttag tttttgtttt tgagtcttgc tatgtggccc 25920
aagctggagt gcagtggcaa gattgtagct cgctgcagcc tcagacacct gggctcgtgt 25980
gattctccta cttcagcctc tcaagtagct aaaactatag gtgcatgcca ctaggcctgg 26040
ctaattttta aaaatttctt gtagagatgg ggggtcttgt taggttactc aggctggttt 26100
caaattcctg gcctcaagtg atcctcccca cttggcctcc taaagtgttg ggattacagg 26160
tgtgagctgc ctcccgaggc ctgtattttt ttatgtatga gtaggtttca ttactttagt 26220
ctggaactct ctacccttct tccatgaaag acacttgatg taatggaata aacactatac 26280
agtctcataa gttgggttta agtcctctcr ctgttatttc cattgtgatc ttgggcaaat 26340
cacttaaatt catctgagcc tttttcttca tctttaatag gtactataaa gctttataca 26400
aatgtattat tatcaaggct cataaaatgt aaacactcag gatatttgat ttagagataa 26460
tagtatgtta cttatgagaa gtagaacata tgagaaatga caagaacaaa ttttcttttt 26520
ggatctagat atgcagatcg tgcaaaacaa attaaatgca atgctgttat caatgaggac 26580
cccaatgcca aactggttcg tgaattaaag gaggaggtga cacggctgaa ggaccttctt 26640
cgtgctcagg gcctgggaga tattattgat agtaagtgaa ttaaggatcg ttacaaaatc 26700
taatcctttc ttcttcaggg ttcttattca gcgttcttat atttaaaata aacttcaagt 26760
taaggagcat gatgaagcta aatggtagtg aaaatgtttt gttttgtttt gttttctttt 26820
tccttaaaaa acccaatgga aatgatctct agattccttt atgtaatgtg taccagacta 26880
tccatttgaa ctcaggaaat tagttttaaa tcgcaaatat gtcactagta tttttgttag 26940
aattctgttt tagatcagca atgaagttta tttggaacgt taaacctgtt atcactgagc 27000
aattaattac ttcatagaaa tgttatttga aattttgagt tttgtttcat gggttgctct 27060
ggattagtgt cgtggcagta atataagaaa cgcaccattc aagcctacct tcagataaga 27120
ttcacctttg aatggccccc tcctcctgta aaatgctgga gtgccccgtg ctgcctatga 27180
acatatgaag gggaaaagtt caattcctat gaagttcaag aactacactt ccagagagta 27240
ttgttttaca atgttgtatt aaaatttatc ttaaaaaatc tccgtgatgt gcctatccag 27300
tggcttattc agaagttctc ttaagtgagt aaaatggatt catataggaa tagaatatgg 27360
tagaatatgg gttagtttag taaagcattt ttttttagtg taaagttaaa cttatattcc 27420
ttgaagttct aataagattt tcaaggcctc ttctagttta attcttctta ttgattagag 27480
tccgagttta gatgaatacc gtttgtaaaa atgataccgt attctaagta ttaacttttc 27540
catctcccca aatgccaaaa actgctctgt tccctctctg tgtcactgct gaacctgaac 27600

tttgaccctt ctcgtatttt ccctcctgct taccttcagt tgatccattg atcgatgatt 27660
actctggaag tggaagcaaa tgtgtgtatt tcacatattg gttatttcca gtactctgac 27720
ttgctaatct gcctctgctg gatcagcctt aaaataagct ttgcatgctg gcaatcctta 27780
cagggaaatc ccagacagag catagtatag aatcagaaat gatcctattt ttgcccttct 27840
tcatttagca aactaatctt tgctgaattt cagtttgtgg cttgagtgaa gttgaaaggc 27900
ttaagaatct cttgctgctt ctactcttca gtttttccca ttgtgattca attaaatcca 27960
aaaatcatag atgatagcaa ctgacactta ctgattttgc ataatatgcc aggccctatt 28020
ctaggcacct tgaatatatt gtcccttta attctgttaa tggctctgaa aatattttat 28080
tatgttcctt ttaccaataa tcatttttag tcagcaaaac tatctatgat ttaagagctc 28140
tgtagatgtt gggctaagac tcatgagtgc ttgtgctaag ttgcttttct ccataacttg 28200
ggcctcggaa tgattagata aacgtttttt ccttttttt ttttttttt ttttttaga 28260
cagtctcact cttgtcaccc aagctggagt gtaatgacat gatcgtggcc cactgtaacc 28320
ttgaactctt gggctcaagt gatctttcca gctcagcctc ctgagtatct gagactacag 28380
ttgttcgctg ccacacctgg ctaatttttt aatttttat ttgtagaggt gaggggtctc 28440
gctgtgtttc ccaggctggt ctggaacccc tgacctcaag cagtcctcct gcctcggcct 28500
tccaaagtgc taggattata gatgtcagcc accacaccca accattcctc ttttcataag 28560
tcacaaatta aaataactta ggtgtctaaa atgactcaaa catttttaaa gagtcacatg 28620
aacctcatta ttctctccac atatatctta agttggtgag gaaattattt cttacccttt 28680
taaggtaatt catcaagaaa ttgttatctt agaagaaaaa tgtcttgatt ttggggctaa 28740
tttttagatt tgtactgagt tatttctttt tttttctttt ttcttttctt ttttgagatg 28800
gagtcttgct ttgtcgctag gctggagttc agtggcggaa tctcggctca ctgcaacttc 28860
cgctgcctgg gttcaagcga tttctcctgc ctcagcctcc cgagtagctg ggactatagg 28920
tgcgcaccac cacgcccagc ttatttttat atttttaata gagatggggt ttcatcatgt 28980
tagccaggat ggtctcaatc tcttgacctc gtgatcctcc tgccttggcc tcccaaagtg 29040
ctggcattac aggtgtgagc catcaagccc agccacctga gttgcttatt tctacagtaa 29100
ggcatgtaat gaggcttggc agacagtgag aaaatgcata ggagaaacca aacatagccc 29160
tggctgtggt tatgaggtta gggcttaaag gcactataat gtcctgatac agaagtgact 29220
cttggactgc aacagctctg tctgagattt ccaaagctca acgatactgg tgtggaaaag 29280
gtgccttctt ctttttgctt ttgtgctcaa ggtgttgctg ctgccagctt gttgcctacc 29340
tcttttctta gcagctgcgg catttcctga aagaccatgc atgactaagt tttaattctg 29400
tattcctgag ggaaagcagc cctgactcct tctcagtttc cattttgatc tatttctctt 29460
tattcctttt agggcttgca tgtgaccatt ggacctaaac tctgtccatc ctggtctttt 29520
ctcccaatgg ctaatttttt gctctaatcc ctatcccccc tttaaaagat gcctctttgc 29580
ttgttgccct atgttctgta tttgcagatg gtggttttca aactgtcttg aagcaaccac 29640
tggctttaag ggagttaaaa aactggcttt aattaaatag tagagatgag taagcccagg 29700
gagaactaat agtatttttt tccagctatg tgtattgact catctccttt gaggaaaact 29760
gacttagttt tgatatgaat tttagtttct caggtttaaa aataagctaa ttttaaagtt 29820
ggattgcacc aatacttaag atttaccttt ggtatctttt tgcaaaggat gactaatgaa 29880
agaaagtaat gaaagaaagt acaaacagca aattaattaa ttttgttgtt tgtggcttaa 29940
ttttttttte ttttaccaat actgagtgta aaagcaaaat taacagttct attaaagact 30000
tagtaaatgg aagtatgttc ctaaaacttg cattttatac taggaagcag gcttagcagt 30060
gacccatggt tgtactgtta ggctaattaa gggccttaaa ttagcaaata atcagcctta 30120
ttttttgctt gctctgacag aagaatggaa cctgcactaa aactcattaa tctgctaatt 30180
gttcttccct gttgaggcaa tcaattagtc ctgagttttt aaccaggact tatacaaaat 30240
tcctctcatt tctataatta ccccttttt atttaagaca tattaaaaat ccttttagca 30300
tgttgaaatg ttgccataca aagcatacta ctcagtgcac aaattttgtt ttaactgtaa 30360
cattttcttg gaaatttttg ctgaattaca cgtaacactg agccagacca atatggatta 30420
ttattgataa ataaatgatc attgttttcc cttactgtta caaactactg gaaaagtgta 30480
aatctgaaga cattgtagac taggaacact agaaagacta aatggaaggt agtttggctt 30540
ttttagggag agtgcctaag cagtatcctt tgaggctctg gtttatcccc ctgcgagata 30600
cagcgttaac ctacattgga atcttttagt ttgtaaatgc agaaatgatt ctgaacctgt 30660
aggataccac tattaaagac agacatgttt tagaactgcc atctagaaca gttgtaacct 30720
gtgtttgtat tatagcgtgt cttggtcttg gcgctttaat gctgtccttt cttacaacga 30780
tattccttcc tgtcttttt tcttcctctg cttctttccc ttttccctac ttttcctgcc 30840
ttctcttctt tctatctccc agatctgaaa gattttcaga acaataagca tagatacttg 30900
ctagcctctg agaatcaacg ccctggccat ttttccacag catccatggg gtccctcact 30960
tcatccccat cttcctgctc actcagtagt caggtgggct tgacgtctgt gaccagtatt 31020
caagagagga tcatgtctac acctggagga gaggaagcta ttgaacgttt aaaggtaagt 31080
aatagttcag actgaataca aggtattcta trtagctcca caaggaagaa ctaggagtaa 31140
aaatcactaa gatttcgact cagcatatgg agaactcttt gactcttaga agtgtcctga 31200
aaattgaatt ttgtgctttg taagttaatt tcttttcatt agaatgcctg agtctgacat 31260
ggccaggcag agttgggaga agtgcatggg cagttcatag gctggcaagg cagagtaaca 31320
gattaataga tgtgtacgtt aattctggga tagtacatca agttacagtg taattgtttt 31380

gttagaaatt tatttggaaa cgtggtttca ttgaaagatg gaattctgtg tagagccatg 31440
cttctgttaa taaagatgaa gtttgggccg ggtgcagtgg ctcacgcctg taatcccagc 31500
actttgggag gccgaggcgg ggtggatcac gaggtcagga gatcaagacc atcctggcta 31560
acacggtgag accctgtctc tactaaaaat acaaaaaatt agccgggcgt ggtggcgggc 31620
acctgtagtc ccagctactc ctgaggctga ggcaggagaa tggcatgaac ccaggaggcg 31680
gagcttgcag tgagccgaga tcgtgccact gcattccagc ctgggggaca cagtgagact 31740
ccgtctcaaa aaaaaaaaaa aaaaaaaaaa aagataaagt tagaacattt gagtagctgt 31800
tccaggttta ttcagatagt tattaaatta tatcttctag aagttcacag aaagccttga 31860
tgccctatga ttactgtaca ggaaggttat ctcctcctat gctggtaccc atggggtcac 31920
agggccatag tgctgcatgt atgcagatta ctggcctatg gctcccagct ttgctgttgg 31980
aggtatttcc cagccccttg ctggctctgt caggttcgat cttatctctt tgtatgaata 32040
tgaaagttaa gagcagcaga gccttctgac tcatagctgc ttctcttcct cctcaattaa 32100
ttcagtgcac agacagttgt cacctgcagt gctgaaagct gttcaggaat cctgccaccc 32160
atcagatgac tgtgacagca tcatctgctt tgaagcaaat atttttttctt tccctacttg 32220
tcacctccag gaagatagtt tgacccatac cttttagcca ttgtgttcaa gagtgctagg 32280
gtaggggaca cagggaggac aaaaaaacact ttttaagagt ctgagaagtt gcagtcatgc 32340
tgccagccag cagtcttctt cacagtcact tggtattccc cttgactgcc attcatgcaa 32400
ggacgataga gaacagtgac acattgacac atttaatagg taaaaggctt gccgaaaaac 32460
tctttaacac taaagtgaat taagcttgat aattgaatat gtcagatgct ttcattttaa 32520
ttttcatgtg aggctgaaaa agattttata tgaagaaaga ggttctgctt tttctttctt 32580
aatcagccta ccttttccct tgacatttta cttacttcaa aagtacagtg tgggccgggc 32640
atggtggctc acacctgtaa acccagtact ttgggaggct gaggcaggag gatcatttta 32700
gctcaggagt ttaaggctgc agtgagctgt gattgtactg ctgcactcca gcctgagtga 32760
cagagcaaga ctctatctca aaaaaacatg ctttagaaac aaaaataagg ccgggcacgg 32820
tagctcacac ctgtaattcc agcactttgg gaggccaagg tgggcggatc atctgaggtc 32880
aggagttcaa gacccgcttg gccaacagtg aaaccccatc tctactaaaa atacaaaaaa 32940
attaagccgg gcttagtggc tcatgcctgt aatcccagca ctttgggagg ctgaggcagg 33000
cagatcactt aaggtcagga gttcaagacc agcctggcca acatggtgaa accccatctc 33060
aactaaaaat aagaaaatta gccaagcatg gtggcacacg cctgtaatcc cagctgcttg 33120
gagggtgagg cacaagaatc acttgaactt gggaggcgga ggttgcacag cctggagatt 33180
gcgccattgc attccagtct gggcggcaga gtgagactct gtctcaaaaa ataaataaat 33240
aaaaatagaa aaattagccg gccatggtga cgtgcatcta tagtcccagc cactcgggaa 33300
gctgaggcag gagaatccct tgaacccagg aggcggaggt tacagtgagc taagattgcg 33360
ccaatgccct ccagcctggg caacagagtg agactccatc tcaaaaaaaa aacaaaaaca 33420
aaaataatgt gcttgtttct tccttcagaa catgtaagtg gaaagtgagt gaagttctgc 33480
aacacctagc aaagggagaa ggagtgagaa atctttccat taaatgttgt aggctgcagg 33540
gtaactttta ctgttggtaa aattgacttt gaaattttac taaagacctt agtgaagagg 33600
agtggtaaga ctccattcac acattttctc acaataggcg tcagtgacca tgttgctttg 33660
cagggctctt ttcagagcca gagtgagaaa tagggagctc cttgggaatg agaagtaaaa 33720
ctcccactta ctgagtcctt aatataaaat gttctttttc tttggggatc ttaagtctta 33780
actatttgct tattaaagcc ttagctcwta cacttgagat tacccttcc tttgaccttc 33840
ctcttgatgt aaaggagaaa tttctcccta actgccaagt gggaccaatt gaaaataatg 33900
tcacatgtta ggagtcaaga cattaaccag actcttcaag attcaggtaa cccagtgtct 33960
tccaaataat ttctagggaa taacactgtc attattcagg gaactctcct tccgacttcc 34020
ctatgtccca ctaaagcaag acaaaactgt attacttcaa tcagaaaaca aatttctctg 34080
ttttaattct tccttacgcc aaaataaaga actcttctgt tttccctagt tttattctag 34140
aaagaaaaca gcttccttag tgttttgtgt tgtcttgagt ttttaatgca gtgtggtagt 34200
tcactgaagc ataagcacct tagttctgct cttatttttgt gaatggtgag aattyctttc 34260
gggcagacct gcattttaat gctgcacgca gataacatct ttcatccgta gcgccagggc 34320
aatggttgtg cagtagaaat gcttgagaat atctaggaac tgggttccta aggcatggtg 34380
ctgcaggaga aaaatattat aatggaaaac agcaatagag gaggttagtt tattttccta 34440
acattttatt tccctttgtt tctttttgac tttaaaaaag gtcagttgat tggatactgt 34500
ttttacacat gatgttgctg tggctttgtt cactgaatag ccttttttcca tatgttaggt 34560
actcatgtca caccattgta aagatgtttt tttttttga gacagagtct tgctctgttg 34620
cccaggctgg agtgcagtga cgcgatctcg gctcactgca acctccgcct cccaggttca 34680
agcaactctc ctgcctcagc cttctgagaa gctgggatta caggcatgta ccaccacatc 34740
tggctaattt ttgtattttt agtagagacg gggtttcgcc atgttggcca ggctggtctc 34800
aaacttctga ctacaggtga tctgcctgcc tcggcctccc aaagtgctgg gattacaagc 34860
atgagccacc atgccaggtc ttcctgtcaa ataaggtttt aaacatgcag ctgtagccac 34920
agctctcgtg tttattagtc tggcatcaca gtcagtactg tgagtgccgt tgcttaggct 34980
tagcctataa atatccacaa atatccatta cagaagcagg caagggcaaa tgctgtgcta 35040
agcagcccaa gggccacatt tggaacccac ttaggaagca tgctatgttc agtacagtaa 35100
aaccttggtg attcaatggc ttcaacaaat gttctgtgta gtaagtgttt ttcatggaaa 35160

tatttctgac ttgtcctggt gatttctgtg tttccttaac tgaatcatca tcaagaagag 35220
attaaattgt ttagctatat gaacataatt tattaaccag tgctatacaa ataactagag 35280
atgcagacag atrtgacgtg agaccaaaga gtgaatgaca cctgagcctg gcattcctcg 35340
gagcttaaag aagagttcct ggtgtccctc tgtttttttgg aatttttttt ttttcaaaga 35400
aagggaggag gattaaagat ttatctcctc aacccctcat ttctgaacca tgaatccaca 35460
tttcaagttc atcttacatt aggcccctag agtggggtat gggaagccct ggacaccagc 35520
atgcctgcct gccagcagaa tgagttgatg ctcctgagac tgaatgctgt tttgcacttg 35580
gcttccctat ttatactaag tttgctactc aggcttgaac agaatatcct tcacttttttc 35640
ttaacttagg gtgaactgtt tcatagaaca tttacttgaa agcctattga tcctgtttta 35700
tttgaagaaa gaacagtgag tcagcaaata gctctagacc tgtctcagtg agtagtgtgg 35760
gctcacttag cttctctggt ctgttggaac acagtggcac catcttggct cagtgcaacc 35820
tctgccttcc aggttcaagc gaatctcctg cttcagcctc ccaagtaact gggactacag 35880
gtgcgtgcca ctgtgcccgg ctaattttcg tatttttagt agagatgggg tttcaccatg 35940
tgagccagct gatcttgaac tcttgtctct ggtccagttt tcttatctcg tgaatggttc 36000
cagaataaac gattttgcag gctttttttca gccctgaagt tgtgctaatg tgactcacag 36060
ttattatttt ctctaagttt cccccaaagt ttttaaaaat agctttctag acaacctaaa 36120
aattgcagct gaaaactgta ataaaaataa aagttgtacc ttactcatta aatctctttt 36180
ctgcaaagag atcatgttct gatttaccct tcagttggta agtaaaatta gataagtcat 36240
attagagttg gctgcaaaat ggtaagagga ttcctagaca gtaattttgc tttcttagag 36300
tgagttctac agttgcccca aaataggatg gctttaatat gactagataa tcctctgctc 36360
tgaaataacc tttttctgaa ggtaatcaac taggaagaaa tatttttaaa atatctcata 36420
tgtcagtttg tattccataa tcaacagtgg tttattgtgg ttagtactat aataagtact 36480
agtcaagatg ccttgtttag gtcgggtgca tcctgttgtc ctgtctgttg aaggttttca 36540
ctttgagaag attgtataga tttctaccca gctatcttta atagacctta aaattatgta 36600
ggttatatca tcagcttgtc ttgtatttgc ctatttaaaa tcatatgcag ggtggggctg 36660
ggaggaaaag aaaaaaaagg aaaaaagcag tcatatgcaa aagcattttc ttttttaaa 36720
gaggggctca ctgtgttgct caggctggag tgtagtggtg cagtcatagc tcacttatgc 36780
agaagcgttt aacaccactc ttagtggttg ttgaataact tttgtggtgc tttatagata 36840
aactattgaa aacttggttc ttggccttct cactccttag ggaattgtgt ggatttgtct 36900
tgctcgctgg ccctctagta tgacactttt gttctccatt tgtggggcgg catgtgcact 36960
gcatggtaat catttctgct gaggactgtg tagagatgcc attcaggtag ctgttttagc 37020
ctttgagcct ctgtgtgtgt tctgaagcct aatgagagcc atttcaggtg gtttttgtca 37080
cttgccaccc ccaactgcag acttgttcta gcttcttcca ccttgtaagt ataactcttt 37140
tctgaaaaag tacctcctac atggctgcct tactgatgcc caggttcctt ctttaatctg 37200
ttttgctctt tgcctacctt tgctggttgt tgttgttctt catctttctc ttttttttaa 37260
attttatttt aaatcatacc ctcctatttg ttccaccaaa ctcatatact actttgcttt 37320
tgcagctgct gctttgaatc tttactcttg ctttttttcac gtgttccctt gccttgcctt 37380
tggtatgttt tttgttttga ggtgacatcc agttcatgta atattctgga agtccaagag 37440
atccagaaag tcaaacacta gacaacttaa ggtttattac tcacaagtcc tttctgtgta 37500
gatattttac gttttggggt cacaggcttc gttagtaacc tgttatcata tgaatcatac 37560
attgaagtgt ttgatgaact ataacacttt tccttgttaa tcactactag catcattact 37620
atatttcatt atcatcatct gactttctta cttcatctcc atgaccagca ttttcacctg 37680
ttttggcatg tgctcgtccc ccttttttgt actccttttc cttccattct ttgcctaatt 37740
gctacttgtc cttcatttct ctgtctgggc ctcactcagc cttgattcac ccatgttaca 37800
gagacctgtg aatactccag ttctagctct tagtacactg ttctcttatc tcctactgta 37860
ttcgttctca gtcctcagac atgctgctct agcaacacat tattcctaga aaactccagg 37920
tactttccca ctctcagcat acccacatgg cagcttctgc tcttctttga tctctgctca 37980
catatcactg agtctgtgag gcctgtcctg acaacctcac tttagattgc tccctcacct 38040
acccttcccc atgtgttctt tcctcttgtt gctttatttt tgtctggagc acttgccact 38100
atctgacata ctgtgtactt attgattaaa atggattata tcatggcggg gcacggtggc 38160
tcacacctgt aatcccagca ctttgggagg ctgaagcggg ccgatcacct gaggtcagga 38220
gttcaagacc agcctgacca atatggtgaa acctcatatc tactaaaaat acaaaaactt 38280
gccggacata gtggtgtgca cctgcagtgc cagctactca ggaggctggg acaggagaat 38340
cgcttgaacc tgggaggtgg aagttgcagt gagccgagat cgcaccactg tacgctagcc 38400
tggatgacag aatgagactc tgtctcaaac agacaaacaa acaaacaaac aaaataaata 38460
ataggattat atcagtctcc ttactacaat gtaaatctta tgaaggcaga gattttgttc 38520
tctgacatat ccataggtgc cttgaagaag gcttgatatg taatggggtt gtatataagt 38580
attgaatgaa taaatcatca cttacttctc tctctctctc tctggtatac agtaacctcc 38640
ttgagagcag gaaacatgtt ttcgttcttt ttaaattaaa ttaaatttaa tttaattaat 38700
ttatttttt aattttactt tacgttccag gatgcatgtg cagaatgtgt aggcttgtta 38760
cataggaata catgtgccat ggtggtttgc tgcccctatc aacccgttat ctgggtttta 38820
aaccctgtat gcattaggtt cttatattta tataagtgta aacactcagc aagtgattgg 38880
taaattgtca cttcagtgtg tgtcttcaga tgcgatttaa tagtaaacct ttctatgttg 38940

actgctgggt tgtgttcagg aggttagtgt gttatggcga aattgtagtg aaacaaaagc 39000
agtaaatacg tttgggaact ttattctttt ttttcttttt ttgagatgga gttttgctct 39060
gttgctcagg ctgcagtgca gtggcgcgat ctcggctcgc tgcaagctct gcctcccggg 39120
ttcatgccat tctccttcct cagcctcctg agtagctggg attacaggcg cctgccacca 39180
tgcccacgtg gatcacgagg tcaagagctt gagaccatcc tggccaacat ggtgaaaccc 39240
cgtctctact aaaaatacaa aaattaactg ggcgtggtgg cacgtgcctg tagtcccagc 39300
tgctcgggag gctgaggcag gagaatcact tgatccccag aggtgaggtt gcagtgatct 39360
gagatcgcgc cactgcagcc tggtgacaaa ggaagactcc gtctcaaaaa aaaaaaagac 39420
ttggaggatt tatttgtttg gggttgattt agttttaggt ctaatcttta tgattcttat 39480
cttatgtatt aataaaagaa aaatctcttt acatttaatt ctggcctgta tcctcaaagg 39540
gcatgatttg cttcttgctt ttctagcatg gtatgttaaa ttataagact ctttgcctct 39600
ttaactgtgc gcttccttcc ttaggaatca gagaagatca ttgctgagtt gaatgaaact 39660
tgggaagaga agcttcgtaa aacagaggcc atcagaatgg agaggtcagg aggttaaaat 39720
ctggaaatgt ttctaagttt tctaggggat gtggattact ttatcgtaag aaaagataaa 39780
atataaatta aaatcaaaga gtctgccctt ctaaaacaca aaacgaaacc aaaaaaatgg 39840
acagggattt atggccagtt gaatttatgc tttgaaatga catctacata tttaaatatt 39900
atctagggag tcacaggttt attagttacc atattacatg ttgctaggga ttcataaaaa 39960
tggaacatat tggccaggcg cggtggctca cgcctgtaat gccagcactt tgggaggctg 40020
aggtgggcag atcacgaggt cgggagatcg agaccatcct ggctagcatg gtgaaacccc 40080
atctctacta aaaatacaaa acaaattagc cgggcatggt ggtgggcacc tgtagtccca 40140
gctactcggg aggctgaggc aggagaatgg cgtgaacctg ggaggcagag tttgcagtga 40200
gccgagatgg tgccactgca ctgcagcctg ggtgacagca agactctgtc tcaaaaaaaa 40260
aaaaaaggaa tagcctcttc ttttcttttt tatattaata tgcatatcat ttatgaatta 40320
ttgactatcc caaagtcagt tgagacaaag ttcttatctc agtgaatata actttaacaa 40380
agcaagattt cattgatttt gaatgtgatg ttgcttttag aaataacttc tcattaatga 40440
tgagataagc acattatttt atgtttgcat tattggctat aaacatggcc agttgttttt 40500
gctttgcagg ctttattttc caattcatat tagactgatt taattgacat tttggtatta 40560
gtgttctgat atacctgttt ttttcctaga gaggctttgt tggctgagat gggagttgcc 40620
attcgggaag atggaggaac cctaggggtt ttctcaccta aaaaggtagg aaacaatgct 40680
gtgaaaccta atcagacaga gacacttttt gtttgtcttt gttactgggg cacttcatgt 40740
ctttcaggac ctactctccc ccttattatc ttgatacttg ccttaatttc agatatttct 40800
gatttcatct tagagcatgg tcttatagta caaaggcagc ctatttctaa aatgcctatc 40860
ttgtttactg tttacaaaag aaaaaaccta aggcaatatg taacctttgg ccatgattca 40920
gtttgtttag tcatctttat ttgttgcttt ctcttaaatt cccagcctaa cctagcgtac 40980
tttcagtttt ggctactctt tcatgtagat tttcccacag aagtcgtcta ccctgcttta 41040
taacttttttg gaaccaattt cagctgtaca ggtttgtgga gaataatact gcttccattt 41100
gcttctgcaa tttgtattat agacagccat tgtactaagt aaagcaattc ctgtccctat 41160
ccttctagac tcctctgctt ccttcttcat tttttaaaaa gtgtctccaa tagttacctt 41220
ccttttttaca taatcatagt tatgaagtag ttacgtagct tttagctaat tagagtagta 41280
agggtttttt tgttttttttc ttaaactaag ttgcaacgtc ttgttggata ctagaaaatg 41340
actatcaaat ggttattttg tgtgcttatg tatcagaaaa taggagaaca tatatcgatg 41400
tacaaagcat gtattttatg ccactagtaa attgtccttt atttcaagtg acctcacatt 41460
ggctattgga gacttaaaca tctttattca taggttctgt tgatgatctc tctaagttag 41520
gatacaaggg gtggtgagtt ttgaattggt acttttatgt gatatgaaat attcctgtgg 41580
ttttcgctac aattcttctg gtcaggtttt tctatcctgg ccacataact ttttttttt 41640
ttttcttccc aaattccaat gtgttatttt caacagaaaa gactgatgta aagacatttt 41700
agatgcatat tgtagaatag aacagttaag tattttatat gtttatatta tgaaaggtta 41760
gtggttttca cagttttaat agtttttcct tcagtaacac atgtattcat tcagatatgt 41820
tacttaggat tatatagatt ttagcagtta cagttttggt ggaatttaat tgtgcctatg 41880
catatgtcta ataaaggtat tttataacat cttacagtat taaagactgt gaggaaattt 41940
ttttttttt ttttttttga gacagagtct cactctgtcg ccaggctgga ctgcagtggt 42000
gtgatcttgg ctcactgcaa cctccgcctc ccgggttcaa gtgattctcc tgcctcagcc 42060
tcccaagtag ctggactgca ggcacatgcc accacgccca gctaagtttt gtatttttag 42120
tagagacggt ttcaccatgt tggccaaggg tggtcttgat ctcttgactt tgtgatccac 42180
ccacttcagc ctctcgaagt gctgggatta caggcatgag ccactgtgcc tggctggaaa 42240
atttttatta taacaactag tgaaaatatt atgcatttaa cataaatcat aaataacttg 42300
aacaaatgtt gagatgtaca ctgcatggaa ttccaaagtg agatctctgc ttccaaagct 42360
ttttcttact ttaattttac ccatcttgct ttagctttct agacttcttc atgaaaggaa 42420
gggatatacc agtgaccaca tccaaactca gtcttaggac ctgtttctga gattccattt 42480
ttcaaattat gatgtgcttt aaagaatgct actcacaaag ggctttgact ttctcaaacc 42540
agtcaagttc ttcaactaca aaggggatg ctagttgaga agatgtggag gtttttcaga 42600
tttctcttta accatctttt atagttcctt ttcctacccc ctcatctgac tcttctttca 42660
ttcttccaca tttgctggca ctctcttgag tagattttgg cctcaagacc atgtgactta 42720

ttttggatt ccagtgggtt attttcacta aagaaggttg gtttttaaa aagattttag 42780
aaatattaga gttagatgaa gaggattgtt agaagatgtt tatcttatcg aatatcagta 42840
attttacttt tttctttaag gtaaagacta ataaattaaa atcttttaag gtaaagacaa 42900
attaataact gggatcagaa gaactgataa aaattatagt gggaagtggt ggggggtat 42960
ggggacattt taataaaagt aaatatttct tgagattttt acccaggcat ttacaaaaat 43020
tataattatt ccgttcagag aattggcctg tcatttttc cttgtgcttt gaccttatca 43080
aaaattaggt cttgtttttt tatcccagaa ctcatgtttt aaaatggaat gccaggcaca 43140
gtggctcact cctatagtcc caacactttg ggaggccaag gagggtgggt cagttgaggt 43200
cagtagttcg agaccagcct ggccaacatg gtgaaaccct gcctccacta aaaatagaaa 43260
aattagccag gcgtggtggc gaacacctgt aatcccagct acttgagagg ctgaggcagg 43320
agaatcactt gaacctggaa ggcagaggtt gcagtgagtc aaaatcatgc cgctgtactc 43380
tagacagagc gagactctgt ctcaaattaa ttaattaatt aaaataaaat agaatgcact 43440
tggttccgtt tggtgaggcc tgaagagagt gttttctctt tattatggaa ccgcaactag 43500
ggataaaaag aaatccctct ctaggctctg tggagagttg gagaagcctg gcctcttgta 43560
ggcccctgtt gttaccacag ctctctcatg gtgccctgct ccaaattgat catctgtaat 43620
ctttttcaga ccccacatct tgttaacctc aatgaagacc cactaatgtc tgagtgccta 43680
ctttattaca tcaaagatgg aattacaagg tatatttatt tcctgttttg gtcacttcgt 43740
gtgttttccc cctcttagat aattgaataa ctaaagggaa ggggttgaaa aaattaacgt 43800
aatgatttgc tgtatttttt gtctgaaata gttacaaact atgctctctt tccaaataat 43860
gtgtttttgc cactggagcc agttactatg tagtttttct ctgaagaccc taaataattt 43920
ttttttcctt taacaaatat acattcctta gggattttat ttgactcatg tctttataat 43980
actgtatgag ttacattggt atatcagtgc tccttgtttc ttattctgat gtagagaggt 44040
tgatcatatg tgaggataga attcatagtt gaaagcttct gatgcaacaa ggcaatacct 44100
tttttgtacg taccaaagga tattctttgg aataaagagg ttcattgttt gtagtgcttt 44160
ctgtgtctga atttccctgg gaaacacttt ctcttgtgtt cagggttggc caagcagatg 44220
ctgagcggcg ccaggacata gtgctgagcg gggctcacat taaagaagag cattgtatct 44280
tccggagtga gagaagcaac agcggggaag gtgagcattc ctggctggag cttcagcaac 44340
aacattttca ttttatatta tgagaaatcc ttaagacttt gtattctctg tctatcagta 44400
gtactttctt atacaatcta attctgaaaa atggagagac ctgggctgct tatgaatgca 44460
gagatggaca aggctgcttt acatgaaaat agcttggaca aaagaagccc ttttttactg 44520
ccaagaactg agaaggacat aggcaattag gcttggtctg gaatgttaat tatttaatag 44580
aaaagtaaga aaatagcaga tatcctgggt aataggagat ttgaaggaca ttaagtcaac 44640
ccagcagaat ttatttttat ctaaaaggga agaaaaagtc agtcttgatt tttgcctggg 44700
ttattaacaa aacaacaatt taatggtttt tttctgttat ataagtcact cattccctta 44760
tcaaaatatt agcttctcag tctttagttt ctggttatta cctatatctc atccttacaa 44820
tttctgatgg ttctgagttt tattgactga accgtcagag atccctgaaa ctaatatttc 44880
ctatcatctt cttaggttta tcaaatagag ttaaatgttc ttgtgttagc catggccaca 44940
gatagcctct tctattgggg ctagttctgg taccccaaaa tgaactactg tatagacaac 45000
ttcagccact tgattgattg cagggattta ttctacttac tgcaaatcct gataagcaac 45060
tgctttccat tatttgattc caatagtttg taatgataac attagtttgt gtttgttcct 45120
cttagttatc gtgaccttag agccctgtga gcgctcagaa acctacgtaa atggcaagag 45180
ggtgtcccag cctgttcagc tgcgctcagg tgagactggg agaggtttgc catcttcagc 45240
aatgtgcaca tggcttctgt gacaactcta atttttggct gtttaaaggc tgaagtaata 45300
gtcagcatta ggatttttgt tcttgtaaaa acaacagctc tgaaagctgt cttttcacat 45360
tagggagggg tgaggttgtt aaatagatac tatatattaa aaaaattatt tcttaaccct 45420
atttttctgt tttgtgctag gaaaccgtat catcatgggt aaaaaccatg ttttccgctt 45480
taaccacccg gaacaagcac gagctgagcg agagaagact ccttctgctg agacccccctc 45540
tgagcctgtg gactggacat ttgcccagag ggagcttctg gaaaaacaag gaattgatat 45600
gaaacaagag atggagaaaa ggtaatgcac agttacgcag cccatatgac tgtttcttct 45660
tttaaacatg taatactaat agcattcttg aatttttttt ttttttttta cttctaggct 45720
acaggaaatg gagatcttat acaaaaagga gaaggaagaa gcagatcttc ttttggagca 45780
gcagagactg gtaggagtcc tgaatctgct aaactgttgg gaaaagggca gcttgttccc 45840
atactttccc tgttccacag agcagtactc acccaaattg cttctgtctc artgatacca 45900
agcactattc tttaatttcc ttaatggaga atgaacttaa atctcccggt agccttagcc 45960
tgaaaaaata gtccacagag gtactctttt gggcttttta tgtcttaaag ccaaatctta 46020
acttctgtta caaccaaata ctttttaagg aaatagagct tttctggtag cctttgcctc 46080
ttgatagtgg ttttggaatt tgcttcagtg gtggttcttt aaatgataat tactctgaat 46140
attgaatttg gtgagagttt gccttggttt tgtttctgat cacttgatag tactaattct 46200
ctgctcttgg gctgactttg ggattgttct tacgctgggc agacttttttt tttttaagtt 46260
aaactgtgtc taaaagtgtt gctgcacagt tgcatgtgtt actcctttcc ttatcccctg 46320
catggagtct gaattctcaa tcaggttctc agtggcatgt gtggtagcgg tgggagcaaa 46380
ggctgcatac ccagcccgga caggacagaa ggcttgcttc ctttgaggga aggaggattt 46440
gagtgagcag ctggaaagtc tgtttaaggt cccagctatt gatacaatac taatggcttc 46500

agcgttttat aatggtaatt gatttcccag atttgataga aactgaatct catgaataaa 46560
taagttgatc cattagttag attttcattt ttaaagccaa ccaaacaacc aattttggtt 46620
aacaaagctg tggttataag aaccttctca gcattcattt gatatgtaag aataaagtat 46680
aactaacctt tggatactgt ggatgcagtt taggaagttc tcatttttta agattcagtt 46740
gtatctttct tttaaggtaa gggggcatta gggagaaagt tcaagcactg gcttctttcc 46800
ttagtataaa atacagcatg taaactacag ccttctagtc atgctctgtg ctttggagaa 46860
gaaggggctt agtgctaagt gggccagaaa cgtgaaggac ttgagcacaa gcaacctttt 46920
ccagaatacc tcagttctca ggcatgggca tggattaaac cagttagttt gttatcttgt 46980
aaattatagt aaagccctaa aaagaatgga ttgtaagaat aaatcaaggg tttggggcaa 47040
gataatagca gtgttaccaa cagaggctct ttaggaggcc tttttagcat taagttttaa 47100
tgaaaacgaa agaactagct acatgtgatt agtggcttga aaggttttgt gacacaaaaa 47160
tgtcttttat cttaataatg tagtacattc taataacatg cctctgatgg tctgtgagag 47220
atcaaatgag gacaggtgta attaagctct attattaaca ataatgcctt gtatttgtat 47280
agtgctttat aatttacaga gagcgttaac atccatcaat ttatttgacc ctcaaaacaa 47340
ccctctttgt ttacttatgc atatgtataa gtaaaatatg taatattacc tccagttgag 47400
gaaaccaaaa cttggctggg cgcggtgggt cacgcctgta atcccagcac tttgggaagc 47460
ctaggcaggt ggatcctggg acctcactta gaataactaa gcacggtgct cgcttcggca 47520
gcacatatac taaaattgga acgatacaga gaagattagc atggcccctg cgcaaggatg 47580
acacgcaaat tcgtgaagcg ttccatattt aaaaaaaaaa aaaaagaaac aaaaaccaca 47640
aaaacaaaac aaaacaagac aaacaaaaaa aaagaataac taagcacggc agaataagga 47700
tgcctggtgc tatccataac caagagttgg tgatcattct gtgccctaaa ataaagatgg 47760
ccactaaata aagagaaagt gaatgtctaa gacgtattta gtttctaggg aatgtacata 47820
ccctgcagat aatcagattc tggttgttgt ttgctctgtg tatctgaaag caagccagga 47880
tttcaagagc agcttcttaa agcaaggaag ttgctttcct ctcaaaggtc tgtctgttcc 47940
actttcattt cttgacttaa gggaggaact gatttctaac acttcagcct gaagaatatc 48000
taccagtagt aatactgaat ggaaaatgtc tttatatatt accctgtatt ggtcaattaa 48060
cagatgtatt tctaatgcac cctgtgaata aacattgaga gaatacaaag gaattattaa 48120
tcagtaaata tttatttaat atttaccata tgtaatcact ttttgcataa tgtcagcatt 48180
aaacaatgga attcaggctc atgagttttc tattctgata gttcccaaat gcatcacatt 48240
gttagggtgc tgagtatttt cttcagttag tgctctagga taccctggtc ctttctgtgc 48300
ttgcttttta cacacagttt tgggatcttc tctcttctta tttgcataag ttccttgcag 48360
agtctcttta ggaggcctac tttgcatttg gcatttaatg aaagccgaag gatcattttg 48420
tgacatgtga attgaccaga cctcagtagc ttttccaact ccttaataaa tgaatctaaa 48480
tattttaaaa atctggccag gcgcggtggc tcacgcctgt aatcccagca ctttgggagg 48540
ccagggctgg cagatcactt gaggtcagga gttcgagacc agcctggcca acatagtgaa 48600
accccatctc tactgaaaat acaaaaatta gcaggacgtc ttggtgggca cctgtaatcc 48660
cagctgcttg ggaggctgag gcaggagaat tgcttaaacc tagcaggcgg aggttgcagt 48720
gagccaagat cgcgcaactg cactcaagcc tgggtgataa agctagactc agtgtcaaat 48780
aataataata ataataataa taataataat atagataaat ataaaaaatt ctttttttga 48840
catggagata ggatttatat cacataaact ttgcatacag tttttagtat caccatcaat 48900
atgtgacagg gtgatacatt tgaactttaa agaggacata ctctacagag tcaaaatggc 48960
agtattttaa aaagaaagaa aatgacagct gctagatttt acttgagttg agttcttttg 49020
gatttatata gacttattta tattttattt tagagtttaa agataacttt tgtctcttaa 49080
gcatgatatt aaaacaattc ttagaatact tgtcatttcc tgaattaggt aatattcatt 49140
tcatccttaa gcaatttgaa gaggaaaaga aaaacagcca gaaaaaaagt gatagttgtt 49200
ggttatgtgg gaaagaatca aagataattt gtttatgttc tgtaaaattc agtttataca 49260
tatttgcaaa aaatcctctt caggagtcag acatagtttt ttctctaata tttatgaatg 49320
tcatatttat cacttaaaat tagccattat gattttaaaa actaaaaatg aaaagaaaaa 49380
ttaatgtaga gtctgtgtaa atttagcaaa cttttgttgg taagttatca aataggtggg 49440
agttataaat gggcttagta ctcatttgta ggtcaataga aataatacaa atcagtggca 49500
agatttttaa attcagaggc caaattatta ttattattgt ggaaagttac ttttttgttc 49560
cttagaaagt gactatacag gcaaatcttt ggttttgtca ggtgtgatga tgtgtgcctg 49620
tagtcccggc taaggcagga agatcacttg agcgcaagtg tttgagacca gtttgggtcc 49680
ctgtctcaaa aaacaaagca aacaaagaaa caaacaaaaa caacaaactt tggtctagga 49740
ggcaaggtac tatttatcag tttcaagcaa acaaactctt ctagagatat ttttggtaac 49800
cacagtgaat aaaagattga ctgttagtaa ctattttcat ttactagact gccctgaact 49860
actcatctgt gagttgagta ttttttttt taagtgaaag catgtttatt aagaaaataa 49920
aggaataggc tgggtgcggt ggctcacccc tgtaatccca gcactttggg aggccaaggt 49980
gggtggatca cctgaagtca ggagttcaag accagccggg ccaacatggt gaaaccccat 50040
ctctaccaaa aatgcaaaaa ttaggtaggc gtggtggcag gtgcctgtag tcccagctac 50100
tcgggaggct gaggcaggag aattacttga acccaagagg tggaggttgc agtgagccga 50160
gatcgggcac tccagcctgg gcgacagagc aagattctgt ctcaaaaaaa aaagaaacaa 50220
agaaagaaag taaagcaata agaatggcta cttcataggc agagcagctg agtattttct 50280

tttcccagca acttttatct cccgaataga aattctttct aatcccactt actatttgga 50340
attctaaatc caacattgct ttcccagaga tccatgttat caagatttgc ttgtattagt 50400
taagagcgat tgcttattat ctatacagcc cttgaaatta cattgttgtg aagtcaggct 50460
gctgttttag ggctaagtaa ggttttctga aaaagaggca atggtgcaaa gacagctgga 50520
aagacagagt taaacatacc ttttaaagtg gatcaatttt atgtaaccag agtatgatct 50580
ttctgtttta tgtgtttatt ggaagagttg tcacttaaag tgatattcct caaatagcta 50640
aaaatgtcct gggttttcta gatagaaaat atctctaagt tgtacttatt aaaatgtgac 50700
ttggggaatt ttaaattata ttcagtggaa gtatatttgc ccagctttat ttttgttcgg 50760
gtaatttatt tctcaaaatt ttggaggtga ggtctttatt gatttgtgtg tgtgaaaaac 50820
aagacaaaac acacccacta tgaggatcct ttaaaaaaaa agctttcttg ttatccagat 50880
aacacacttt aatcgtacga cggaaatgga tacatttaac atgacaaaaa atagcaaatc 50940
ctctgctcat ggaaagatga tttccaaacc ctcttgtctt acagaagaaa acatttggtt 51000
gactaagaac tgacatggtc tgctcatagg gattcttagc atttttaaca tcatccctgt 51060
ttgactgctt ctttttctac attctcatct tctaactttt ttctagggct gctcttttta 51120
gtaaagctaa acaaaacctg gcgtcctaca gccttactaa cctttgttaa tccgcctgct 51180
cgtcacatct ttgtatggca tctgtgctca ctcatggggg aaagagattc tggaagaagc 51240
aacagcaata accaaggcaa atgtggggaa gtgtctgcat gggttgttag catgatgtgt 51300
ggggcctagg ttttacttca ttttcttcga atagtgaggt cgttaagtca cgccatttga 51360
ataattttgt atgaccatag ggtgacatac tcatccctgt tgtcaattct cctaagaaat 51420
ttggtggcaa atttcatcac ttaagccaga aagtggaaat ttggattagg gatgccaaat 51480
aagcaaatct ttatattgat agaaatgtac ataggcgtat gtcttaatag acttccacat 51540
tattgagggg tttttttgg ttttgttttg tttttagtt tttttggtct tatttttaaa 51600
tttggttaag gctttttgc ttgctttttc ttcgatttaa ttttcctttt ttacatttta 51660
attttggtta ttttggggcc attttttgat tttgtttttc caaaggacgc ggattctgat 51720
agcggggacg attctgacaa gaggtcgtgt gaagagagct ggaaactgat tacttctctg 51780
agagaaaagc tacctcccag caagttgcaa accattgtta aaaaatgtgg cctcccaagc 51840
agtgggaaga aacgtgaacc aattaaaatg tatcagatac cccaaagaag gcgcttgagt 51900
aaagattcca agtgggtcac aatctcagat cttaaaattc aggctgtcaa agagatttgc 51960
tatgaggttg ctctcaatga cttcaggcac agtcggcagg agattgaagc cctggccatt 52020
gtcaagatga aggagctttg tgccatgtat ggcaagaaag accccaatga gcgggactcc 52080
tggagggcag tggccaggga cgtctgggat accgtcggtg ttggggatga gaagatcgaa 52140
gacgtcatgg ccactgggaa aggcagcact gatgtagatg acctcaaggt tcatatagac 52200
aagctggaag atattttgca agaagtcaaa aagcaaaata acatgaaaga cgaggagata 52260
aaagtcttaa gaaataaaat gctcaaaatg gaaaaagtct tgccactgat cggatctcag 52320
gaacagaaaa gcccaggaag ccacaaagca aaggagcctg ttggtgctgg tgttagtagc 52380
acctctgaga ataatgtaag taaaggagac aatggagaac ttgcaaaaga agaacgtgtt 52440
tcccagctga tgaatgggga tccagctttt agacgtggac gtctgcgctg gatgaggcaa 52500
gagcaaattc ggtttaagaa cttgcaacag caggagataa caaagcagct tcgtcggcag 52560
aatgtacctc ataggttcat ccctcctgag aaccggaagc cccgcttccc ctttaagagc 52620
aaccctaaac acagaaactc ttggagtcct gggacacata tcatcataac agaagatgag 52680
gttatagagc ttaggattcc aaaagacgat gaagcaagga aagggaataa agaagagagc 52740
caagaaaaag ggggtaaagg agcttttaag gatccccagt ttccatgggg ctctcaagga 52800
atgagaagtc aagatcacat ccaagttagc aagcagcaca ttaataatca gcaacagcca 52860
cctcaactac gttggagaag caattctctc aataatggcc agccgaaaag tacgcgctgc 52920
caggcatctg cctccgcgga gtcattaaac tcccacagtg gtcaccccac tgctgatgta 52980
cagactttcc aggcaaagcg ccatattcat caacaccgtc agtcttactg taattataac 53040
actggaggtc agttagaggg caatgcagcc acttcctatc agaagcagac tgacaaaccc 53100
agccactgta gccagtttgt gacacctccg cggatgagga gacagttctc agcacccaat 53160
ctcaaagctg gtcgagaaac cacagtataa atcagttact ggacaaactt gaaatcatgg 53220
tggaagaaac agacagtgtt agctcatgat ttgatttggt tctacctttg gccttgagtt 53280
cttattattt acattataaa tattaactgg ttttatattg ttaagacaaa acactggtaa 53340
aagtttcaac acctcccttt tgcttgtata ccataaatgg gcagtttctg aaattttgga 53400
taaagcatca agaactcctt tttctgaaac gttcctcctt ttttagtgcc taattaatat 53460
acttacttac acagacttgt cccatcttga tgtaagtttg ttatgttttt ataatgccta 53520
taaattaatc tgacatccag aaagacctgc ctcttagttt atgcagacat tcaaaaggag 53580
aatttgataa gcagaattaa atgtctgtta ttcataaatg ttagaaattg gtttaaaaat 53640
aataacatgc aactgggtgg gcacagcttt ggatttctca tccacgtctg ttcccttctt 53700
taaatatgat tagaattttc gtaagccaaa actaaaggtc ttcaaacata tcttgcaata 53760
catagctttg catcttccac acaacttcca cgtcttcttt tattctgtat tttacaaaaa 53820
actcatataa tggtttttc ttttgaatgt atcagaatat gtatgaaatt tgtcatatca 53880
gaatctataa attggcactg tgtctgggat tctgtttgtg gtttgcattt tgaagttcaa 53940
tgtatcattt gacacacatg cacacaaagt gaactattca ccaaaatcat gccaacagaa 54000
ataccaacaa aaaacttgtg tagcttgtaa ggaggttatt tggcttgaga gggaagaggt 54060

gggacagaca tgggggacaa agctacagca tatggctctg tggtgctgct tttttccaat 54120
agttgatgaa atagtgaaat aaaagttaac tttggttggg cttggaataa aggagatgat 54180
acataataaa ctattctttg gttagcatta gtaatgcttg tagacactca aacaaggaag 54240
gcgaagtcct gatgatgatt tattgtatca atacctcatt ttatttggtt gtggtctcag 54300
tttgcctcac tggagaatcc actaagaaag atggatttta atgggaagaa aagaattatt 54360
ttctacatcg ttcctcatct cttgagtttg tcaaaaagtt ttctaagagg cagtgtgaag 54420
agccagagat gtccgaaatt gcctgacttc atgttttgga agattcctaa gtcacatgaa 54480
atcatctatc aaacacktgt cttccaaacg gtgtgaaaaa ctacagccaa gatagcagtt 54540
agcattcaag gaagccttaa agattgtgat tatgtttttt tttcctttgc gtgtgggcac 54600
tatgctttat tagaacacat tatttttaat catagtattt tttctttgcc tctaagaaaa 54660
tactttctta atgctcagaa agttgcttcc aattaatgtt tttttcccct taaaaagaga 54720
agctttgaga gatatttttg ctttcatagc tagaacagtt gaagtcttca actgaggttt 54780
tatagcagat tagacatggg taaatgatgt ctgtaatggg ttgagttact gagatgacaa 54840
tctcctgtgc catttggttt gaatgtactt gataggctgc ttcaaatcag tcacttcaat 54900
gcattttgtg taaacccagt tgtccttttt tattcctctt tagacataaa tgtgctaact 54960
attcctttct atacacaatt tattttctaa gattaaaaat aataattgct aggtttggtg 55020
gcacatgcgt acagccccag ctacttagga ggctgaagca ggaggatccc tagagcccag 55080
gagttctggg ctgtggtaca ctatgccagt tgggtgtaag tttggcatca atatggtgat 55140
gtccctgaag caggaaacca caaacttgca tcaggacggg tgaacccagc ccaggttgga 55200
aatggagcag gccaaaactc ccaggctgat cagtaatggg atcttgccta tgaatagcca 55260
gtgcactgca gccaggacaa cagagtgaga acctgtcttt aaaaaaaaaa aaaaaaaaaa 55320
aaaaaggtaa tatttattat attattcagg tttcaactct gtagcaaaaa tgggctctca 55380
tttccctaac ttgagataac atagggtagg tccatatatt ttcattctta cagtggtctt 55440
ttcatgggag tgaatgagtt actctccact ggtgattagg taatactgta gaatgaagag 55500
ttgtataata tattcattta cagctgtgga ttgtggtaag gactatgtcc acagtgatat 55560
tccaaagaat tgggtttata tttgtgcttc atctgttaat cccaggtgtc ctcatgttgc 55620
tgaaatattt agatagctaa aatatccctt aatttcacag atgaccagga agaaattaac 55680
caaggtttta ttgactgcca tgtatgtccc atgatgcatt tctgagcaaa tgcttatcct 55740
agagaataac tctgtatgaa taaaattgct taattgagtc tcttactaaa taagtaacta 55800
gtgccatgct tttgtgagct cttggtatgg cccatattac tttgtttttt gtttttgtta 55860
ttgttgtttt gtgatagtct tgctctgtcg cccaggctgc agtgcagtgg cacaatctca 55920
gctcactgca acctctgcct cctgggttca agcaattctc ctgtctcagc ctcctgggta 55980
gctgggacta caggtgcatg ccaccatgcc tggctaactt ttgtattttt agtagagaca 56040
gggtttcacc acgttggtca ggctggtctc gaattcctaa cctcaggtga tccacctgcc 56100
ttggcctccc aaagtgctga gattacaggc gtgagccacc gcgcctggcc tgtttgtttt 56160
tttaacatga tttttctcta agcttaaata ccacaaggcc aaagagaaat ggtcataatt 56220
taaaccatta ttatattgtt gaggtatccc tagctattat tatagcaaag tgggaaaaaa 56280
gtgtttattc tattgaagtt atgtaatgat ccgacattaa tgggaatata gaggagtcct 56340
aattaattgg tataatttca caaagcggaa tggtattcct tggagagtta aagacattct 56400
ctttagtaag tgtaaaccaa agggcatttt ctttattcct gcttctaatt ccttctagcc 56460
cagtgaatta tttctctttt cactggtaat gtgatgaatg ggaattgttt attacattga 56520
agtgacttga agtgaccttt tgtgctttag gtgcaggttg acactgaaaa aaaaacaaaa 56580
crctgaattt ttcacaccta tgtctgcatt aaaggctgtt ttactaccgg aagttacata 56640
gacttcctgc agtcagctga tgtgccccag tgccttactg gtcctttgta gatttgcctt 56700
aatgatttgt acaaatgact gggaggcggg gatgctgcct gtgtcctggt gaaccttaat 56760
gaaggggccg tcttaggcac agtgcaaaac aagcatttgt cctgtactgt tagagccaaa 56820
attgtgatga gcaatactga taattgtcca gtttatgtca tctttcccag attttaaaat 56880
ctgttctaga tattcttagc ttgaaccact tttgattgtg aaatgtatta ggtgttgtcc 56940
cattattact gtaaaatgaa gttttgaatc ttcttgttaa taaactgtgg atttcccctc 57000
tcaatttctt aaacaacaac aaaaaaatgc ttgaagattg tctttgagtg taagatctgc 57060
cttttcagaa agggagtgtt agtttgtaat gttaaaaaat aaagacctca ttcaataaaa 57120
gttgaagtca tcttttaaga gtgtgatttc tctctatgtg ggaagaggga aaaggaaagc 57180
aaggtaatgc taactaaacc tggttttgac ttttatttat ttgcttttc atctgaagat 57240
ggttgtaaac taaatcttct ttttgatatt ttctatgtga acttgattag ttttaaagct 57300
tttgctttag ctacccttt ctgtttaaca aaattctatc ttaagtgaga tctttcagcc 57360
ctattttatg ccacatttca cttacataaa atcttttctc acaaaagatg aataaggata 57420
ttactttctg atttgagttt agtttaacca tgttttcagt tttttctcta catactcttg 57480
gtttgaagtt cttttaaatg aaatgattta ttttccttta tatgggcaga aactaaattt 57540
ctaaaaagtt ttgagtgact ctgcaggttt tctttctcct ttaacctctg cttttcttca 57600
cccagagtat tgatggtaca ctagagagga aacaggacaa agcagaatat ttttctggaa 57660
tggaccacta gtctcattga aaatttggac ctgacattat tttttaatgt ttaaaaacaa 57720
cataatcacg catcatggga tagaatacaa aatccacatg agtttttaat atgaaatctg 57780
aaactttaga gaagtgttat gcctgaatcg gtctgcttcc agcctcatac ccttttcctt 57840

caggtacttt gtaatccttt gctgtcggag aaactgttgg aaaagttgca atgtccctac 57900
cgtttatagg gtgacacaga agattacagt gtgatatgtg tagtgacact gtgctgatga 57960
gactgtggcc acacagtgct gctggtggtg agaagcgaag ttaagtgttg tcaacaactg 58020
gggtgtcatc ttgggtgtgg atattttcct tccagctatt gcttgctaat tataatctgc 58080
cattaggagc caagccttta atgtctaaaa cagcctaacc ggaaaactca gaatggtgat 58140
attgatttta ctgtgaatat caccttctca gaggtgcctt aattgtacca ccatcatctc 58200
tcatcgagag taccctagct tgtctctttg cttctattct tgacccctat aatccaatgt 58260
ccacatagca accaagatca tcttttaaaa acataaatca tatcttgtca gtgtcctcag 58320
gggctttcca ttgcatgaaa gaatcttagt cttactgtag tctccaaggc tttacatgtt 58380
ttggctgctt acctaccact cttaacactg tcactctcac accacagaca ctctgcccct 58440
ttttttgttc cttgtatatg ccaagtttct tcccacctct gggctcttgc acttctcaag 58500
cctggagccc atttccatca tgcttacctc cccttccaga cttgagctcc tcagctcttt 58560
ctgcccagcc caccacaagt tgactcactg ttttaataat gatggcctgt aattatcttt 58620
tctttgtatt tgtttggttc ttccacgaga atgtaaactc acctagaata aatgccgtat 58680
ctgtcttatt cacttcgcca gtcatataaa tccataacct ggaagagtgg ttctcaaact 58740
ttaatgtcag ttagaatttt ggggtcttat gagaacacag actattgggc ccaccttcag 58800
agtttctgat tcactaggtc tggggcgaga cctgataaat ttgcatacct cataaattcc 58860
caggtgattc tgatgctgct ggtccagtaa acacctttga taaccagtga tctggaatgg 58920
gggctctggt aagcattttc tgtaaaacac tagatagtaa atattttaga cttagtggac 58980
catgtggtct tttttgtgaa aactgttcaa ctctattgtt gtagtacgaa agcagccata 59040
gacaatatgg aaatgaacaa gggtggccgt gttccagtca aactgtgttt acagaaacag 59100
gcagtgggct ggatctagcc cgaaggccat agtttgccga ccccgattag aagatagcta 59160
ctttgataaa tatttatcta aaagatcagt atgtgactgt gtatgtaatc tagaagacag 59220
atacatgcca gttgctggga tgagtggtaa gtgacaaatg tgtggtccct ggcctcataa 59280
agttctcaga gttatggagg gagacagctg taataaccac accagtagtg agtcagtaca 59340
tttgtgacaa gtgccatgat ggaaaaacac aaggtgctct gaaagggttt aacgggtgcc 59400
tttgggaatg aaaagtggcc tctcattgct cttagctgac aatgtttttg tttcactgag 59460
gaaaaaaagc aatcagaaga aactttgttt tcttcctttg tctagttttg cctgcatcat 59520
ccatccatac ctgtactggg ctttccctcc tgttagtgga taaactgtcc ttgatcctct 59580
cctagcccag cccttgctcc ctgcattcca tgccctctca agaacttgtt gaacatttac 59640
ccccttctct tttacatcat cagtttctct cctggattat tcctgccagc atgcaaacat 59700
gtcaccagtt gcaaaagttt tctgtactta ttacgtccta cttgcctacc tccattcttt 59760
ttttttttt ttttgagata gagtctcact ctgtcaccca ggctggagtg cagtggcacg 59820
atctgagctt actgcaacct ctaccttccg ggttcaagta attctcctgc ctcctgagta 59880
gctgggacta caggtgaacg ccaccacacc gagaattttt gtatttttag tagagacgga 59940
cggggtctca ccatgttggc taggctggtc ttgaactcct gacctcaggt gatccacctg 60000
tctcagcctc ccaaagtgct gggattacgg gtgtgagcca ccgctccagg ccccatttt 60060
tcttttcatt atttcctatc ctcattcttt gctggaactc atctagtcag actcttgttc 60120
tcccctcacc accaaagctg ctcttatcag agtcacaaaa gctcttcttg atgatagatt 60180
cagtgcattt tccagtcctc atcttactga atgtctctgg catttgacac agttgtctgc 60240
tcttcatggg acactttctt ctcttggctt cttttccttg aacctttcca gctgctccat 60300
catggtctgt ggacacagac cttggctatt tctatrtgtt tctctgttcc cccttccatt 60360
tttgtactga tgatgaagat aatgataact catatattga gcatgtatta catgtattaa 60420
tacacataaa ccttttacat gttttgactc attgagtcct cacaacacyc ctcttctgtg 60480
aggtaggtat tgcttattcc cttttcacat actttcccag tgtaacacgg ctactcagcg 60540
ggagagccag agccagtccc agggagtcat tctgtctgca aacctctgcc ctgcctcgcc 60600
ttctgctgac tctcaggtgt cctctctctc cagcctgggc ccctaccctg agctctagat 60660
gctgatgata cacaactact gcctgacatc caaatcactg gcaggcacct tacatttaaa 60720
tcatttgaac tggagcagct agctcaggcc acaaatccaa gtcaccttta gtttctcttt 60780
ttcttcacat cccatgatca gatcatcagt agatcctgtt tgctctccct gtaaaacata 60840
tcctgactcc atctcttttc accacctttg ctgctaataa cctggtccaa gccaccacag 60900
tgctggacag tgctgtgaca gcctcatgac cattcttctt atttccatag ttcctcctga 60960
gtcccacaga gaataatgtt tttaaaacat atcccagctc taaaccatcc agcagtttcc 61020
attagaataa tatccccaga ccttaccatg gtctataagg ccatctgtac tccatttgct 61080
acctgattac atctgatgca tctcctaccc gtctccctac tctgccattc catccacaca 61140
agttttcttg cttttcctga aatatgtggt gccccctttcc atttcaggac cttggcactt 61200
gctgttacta tgaatttgag tagaataacc cctgccaaag acctctgcct gcttaaccta 61260
ctcagtctag aagtctgtct cccattcttc tccctggtag tctctaaccc tatctttatc 61320
tttcttcata ctagtactcg ccactaaaat ccatctatat tttgtttact ctttatcatc 61380
tctgtctcct ttgtcaatgg aacccttgtc ttattgacca ctagcatttg gcacatagta 61440
ggagcacaaa aggtatttgc tgagtggatg gatggataat ggtttaaaaa tcatgctctg 61500
gattgcctgg gctcaagtct cattttagc atttgcttaa cctctttctt tctctttaac 61560
atgaggctaa ctatatagta ccatctttat ttttttgttt tgttttgttt tttgtttttt 61620

tattgagaca gagtcttgct ctgtcaccta ggctggagtg tgtgacacga tcttggctca 61680
ctgcaacctc tgctgccctg gttcaagcga ttcttgtggc tcagcctccc aagtagctga 61740
gactacaggc acacgccacc acgccctact tatttttttg tatttttagt agagatgagt 61800
tttgccatgg tgaccagact ggtttcaaac tcccgacctc aggtgatccg tccaccttgg 61860
cctcccaaag ttctgggatt acaggcatga gccaccacgc ctagcctagt accatcttta 61920
taaggttatt gagaagattg taataattca tgtttaacat taagaacact tcctgataca 61980
aaataagctt aaaaaccggg gtagataggt aggagagaac attctagtag agagaacagc 62040
cagtgctgtg cctttataaa tatggggcat attttgaact agttaggatt ttttttttt 62100
tttttttga gatggagtct cgctctgtcg cccaggctgg agtgcagtgg cgcaatctcg 62160
gctcactgca agctccgcct cctgggttca caccattctc ctgcctcagg ctcccgagta 62220
gctgggacta caggtgccca ccaccacacc cggctaattt tttgtagttt tagtagagac 62280
ggggtttcac cgtgttagcc aggatggtct tgatctcctg acctcgtgat ctgcccacct 62340
tggcctccca aagtgttgag attacaggcg tgagccaccg tgcttggcca ttaggatttc 62400
ttataaacat aaatgtctac tgaaaaaatt attattatta ttattattac tgagacagag 62460
tttcactctt gttgcccagg ccggagtgca gtggcgagat cttggctcac tgcaacctct 62520
gcctcctggg ttcaagcagt tctcctgctc agccttccga gtagctggga ttacggcatg 62580
tgccaccatg cctagctaat ttttgtatca ttagtagaga tggggtttca ccagtcagcc 62640
aggctggtct cgaactccta actcacccgc ctcggcctcc caaagtgctg ggattgcagg 62700
catgagccac cgcgtccagc caaaatatat tatttttaa tgtgccttcc gtaattagtt 62760
tttgtcacaa aatgaacaga acgatagtat gtgaatctca ttaattagcc atttgttgtg 62820
ggaaatggtg atgatataca gtgtcctttt taaaagctcg taagtagcct ggtttgggga 62880
gggggccaat tcaatagatg atgttgatat tttgactatc aattgtggtc aaaatgcagt 62940
taaagttttt tgtttgtttg tttttgttgt tgtttttgag atggggtctc gctgtgtcct 63000
ccaggctgga gtggagtggc acagtcttgg ctcactgcaa cctccacttc ctgggctcaa 63060
gtgatcctcc caccttagcg tcccgagtaa ctgggactac atatgccctc cactaagtcc 63120
ggctaatttt tgtgttttta gcagaaacag ggtctcacta tgttgcccgg actggtctca 63180
aactcctgag ctcaagtgat ccacccgcct ctgcctccca gagtgctggg attacaggca 63240
tgtgccacca cacctggccc agttaaactt tttagattca tagggctatg ggaaatcttt 63300
tcatgttagt tttggttact aaagcacagc cacagctgtt agggaaggtc atcctcatac 63360
cttaactgtc agagtctgag aattttgtat tcattggtag tttgaccagc ctgtcaactt 63420
attcttgaag aaatcctcaa gataaagatt aacatttttt ttctacctga tgatttttag 63480
gtaaagcagc agctttctct ttttgctata ataattcatg taatgtatag tttttccccg 63540
actccccagc tttttgagac atttgaaaca gagaagctga aaaaaaaact ataatgaact 63600
tttgtatacc catacttata tatttttcct gaattatgtg aaaattatag atatcatggc 63660
agtttacccc aaatacttcc atgcttatct cttaaaatca agaatattct ttttttttt 63720
ttttttttt ttttaagacg gagtctcact ctgttgccca ggctagagca atggtgcaat 63780
ctcagctcac tgcaaactac acctcccggg ttgaagtgat tctcccgcct cagcctccca 63840
agtagctggg actacaggca cctgctatca tgcccagcta attttgtgtt tttgtagaga 63900
tggggtttca ccatgttggc caggctggtc ttgaactcct gacctcaggt gatctgcctg 63960
cctcggcctc ccaaagtgct gggattacag gcatgagcca ctgcatccag ccaggaatat 64020
tctcctaata ccactatccc atctaggaaa tttaatactg attcaatatt gattcaatat 64080
tatatcatgt acagtccatt ttcaaatttc gttaattaaa tcaaaaatta ggaaagagag 64140
tgttttaatc caagatccag tcaaggttta tatatttcat ttggttatga ttttttaaat 64200
ctttcattat actggtattt ttgaattgtc taggccacat tctgatttgt ctgattattt 64260
cctcagtatt agattcaagt taaatctttt tgttaagaat gctataaagg ggccgggtgc 64320
agtggctcac gcctctaatc ccagcgcttt gggaggccga gacgggtgga tcacctgagg 64380
tcaggagttc aagaccagcc tggccaacat ggtaaaaccc cctctctatt aaaaatacaa 64440
aatttaactg ggcatggtgg tacacacctg tagtcccagc tacttgggag gctgaggcag 64500
gagaattgct tgaacctggg aggcagaggt tatagtgagc tgagattgtg ccactgcact 64560
ccagcctggg tgacagagca aaactccatc tcaaaaaaaa aaaaaaaaaa aagctatgaa 64620
ggggccaggc acgttggctc actcttgtag tcccagctac ttgggaggct gaggcaggag 64680
gatcacttga gcctgagagg tcaaggttat agtgagctat gattgtacca ttgcacccca 64740
gcctgggtga cagagggaga ccctttctcc aaaaaaaagg aaaaagattg ctataatggt 64800
gatgttgtgt acttcctgtt gcatcacatc agtagacata aaatgtcagt ttgtcctatt 64860
atcaataatg ttaaagtttg gtaaatttgg ttaatgtgat gtttcccagc tctctgcatt 64920
gtaaatatat catttccctt cttaaataga tatctgggca tggactttgg atcgtatgca 64980
tatcctgttt cccatgcatc ttttactcaa ttattttagc tttttgttgt ttgttgctat 65040
tgttgttgtt tttaagatgg ggtctcttgc tctgtagccc aggctggagt acagtggcac 65100
gatctcggct cactgcaccc tccgcctccc gggctcaagt gatcctccta cctcagcccc 65160
tctagtagct gggactacag gtgtgtgcca tcacacctgg ataatttttg tattcttttg 65220
tagacatggg gtttggccat gttgcccagg ccagtctgga attcctgggc tcaagcagtt 65280
tgcctgcctc agcctcccaa agtgctggga ttataggcat tgagtcactg tgcttggcct 65340
attttagtat ttattatcct ttcttgattc agttattaca tctggggttg caatatggtg 65400

attttccaat tgccatttct tctgcattta ttagccacct ttttttttgtt tttttgtttt 65460
tcagatggga tcttgctgtg ttgcccaggc tggagtggct attcacaggc ctgatcatag 65520
cgcactacag ccttcaactc ctgggctcca gggatcccag gggctgggca tggtggctca 65580
cacctgtaat cccagcactt tgggagataa tttaactttc tcagaattag aaggctagta 65640
agaaaaagtg ccaggacccc aaacccaagc ctctttagcc tcgagtagct gaaatgaaca 65700
ggtgcatgcc ccatgcctgg taattagcta gcttttaaaa aatttttatt taaagaagag 65760
tttccccttt ttctctcctt ctttcagtat caccatggaa tcatgaattc tttctatatt 65820
caatgtactg tgctatttttt ctgttgattt attggccagt gggatcctgt gtccttttga 65880
aatggcccca gtggtgtttg agttcttcca tcctttctga cacaaaatat ttcagactca 65940
ccttgtactt tccttgcttc agacctggaa gcagctattt ttttcaagga accctggttc 66000
ctttcagtgg cagtggtact tagaaaccaa gatctaggtg ctaagtatgc ttaaacaata 66060
acctttttata aaaaaacatt gcctgaggcc aggcgcactg gctcacatct ataatcccag 66120
cactttggga ggccgaggtg ggcggatcac gaggtaagga gttcgagacc agcctggcca 66180
acatagtgaa accctgtctc tactaaaaat acaaaaaatt agccaggcac ggtggcgtgc 66240
acctgtagtc ccagctactt gggaggctaa ggcaggagaa tcgcttgaac ctgggaggca 66300
gaggttgtgg tgagccgaga tcgcatcact gcactccaac ctgggtgaca gagcgagact 66360
ccacctcaaa aaaaaaaaaa attgcctgac attttgatat tgtattacca tgtaagaggg 66420
ctttcacgtg tttttaaaat ggttcatctc ttggtccttc ctctagtact gtgtgggtag 66480
gcaggtcaaa tgctaaaatc tttatttaaa gatgaagaag gtgaagttta gagataattt 66540
aactttctca gaattagaag gctagtaaga acaagtgcca ggaccccaaa cccaagcctc 66600
ttaccagttt aatttttttt tctttttttt cataacatgt agtctcctaa aaaagtgtgt 66660
ttagatatct agtaaagcta tgtagttggt tcgtggtctt ggtcactgcc tgcccatcta 66720
aagagaacac ttcctttgtg agaataactg acaaacgacc agaaccatag aggatatttg 66780
catccttctg agcattgctt gggagctcct tttgtaatgg agtttgctaa atggctttca 66840
gcagtgcttt atgatgcatg gaatgtttga tgtgggactt ggagagctct gtttggaaac 66900
agtggatagt catacttata ccctttgtct catgacagcc tgtaattggg tcactgctag 66960
ggaaccaaaa gaagcagggc tgggtgcggt ggctcacgcc tgtaatccca gcactttggg 67020
aggccaaggt gggcagatcg tgaggtcagg agttcgagat cagcctggcc aacatggtga 67080
aaacccatct ctactaaaat ataaaaatta gtcgggcatg gtggcgggca cctgtagtcc 67140
cagctactga ggaagctgag gcaggagaat cgcttgaacc cgggaggtgg atgttgcagt 67200
gagccaagat tgcgccattg cactccagcc tggacaataa tagagtgaga ctccatctca 67260
aaaaagaagg aggaaaccct ggcagtaaat cttaaaatgc tatgcttatg ctgtaattta 67320
gtattgcgct gtttgaaggt ctctttctta accctactta ttttattttc actaaaattc 67380
aagtttttca aattaacaac ctaaggaaac tttttagaca tttgtcccct attctctgcc 67440
caccatgaga gtttaatacc tcagatatac tacatatctg tatactgatt atatttgtac 67500
ttcatatatt aaaagactaa gatttttttc atcctctgag aaccattttt tggccccttt 67560
ggggtgatat tacccttatt gaaaaacgca tggagtagat ggaaaaatga aggatacatc 67620
tgatttgtat tacaataatc ctttttttttt ttttttttga gacagagtct cgctctgtct 67680
ccaggctgga gtgcagtggc gcgatctctg ttcacttcaa cctctgcctc ccgggttcaa 67740
gtgattctcc tgcctcatcc tcccaagtag ctgggattat aggcacgtgc caccatggcc 67800
agctaatttt tgtattttta gtagagacag ggtttctcca tgttggtcag gctggtctcg 67860
aactcccgac ctcaggtgat cacccacctc ggcctcccaa agtgctggga ttacaggtgt 67920
gagccactgc acccggtcag taatccatct tatagttgca tattgaaaat aaagtgtctt 67980
acttgaataa ctaaatgcag tctataaatt atttcagtcc taatagcact atgagctttt 68040
ggttacctga aggctaaact catcctactg atcaaagata gtaatttttc acaaagagga 68100
agcatcaaag tctacaatct taatttttttt tccttccaaa gagttagata attgagattt 68160
ttgatgatac tggtgaattt aatctttttc aatattgttt ttctctggaa tttattgata 68220
tataagatat tacttgaata tattctaacg cttaaggtat acacgtgtta tttcacctaa 68280
aaattccccc tttctcctac cactacccca gatgaagaac tcttgagtct ttgagggagt 68340
agaaacatac tgcagtgagt tttttctttt taacatttac ccattacaga ccatcttgtg 68400
tgctttgtgg gttttttttcc taaacatttc cctgtatgtc tttttattat tgtttcctgc 68460
ctttgtctct tttccttgtc ctttgctttc tctacttccc ccaatattcc ttccatttct 68520
ctttccctgc cctcttattt cttccctctc ctacatgtta tctcctttct tttcattcag 68580
gactatgaga gtaaattgca ggccttgcag aagcaggttg aaacccgatc tctggctgca 68640
gaaacaactg aagaggagga agaagaggaa gaaggtgaaa tctagagacc gaaagtttcc 68700
tgtgtatatc tttttgaagt tatattatca aattagtcat ttatgcataa tcaaagctgg 68760
attcctctta gttggcccta tcatttattg acattttact gagccagttt acagatgatt 68820
gaatagatgt gcagtgttga gaatccaggc tgtattttat gaggtgggga tgggagtgat 68880
ggtgttctta gaccttcgca cggtttctga gcacgttcat ttaaatccct taggaagttg 68940
acctgcactg tctctcatat tgctgtagat gtgactttta gatatccatt acaaatatag 69000
agctgaatct ttgttaagac agtctgttca atttgaagcc atttctggca aatttctaat 69060
catattgctt agtcatgttt gcctatctta ggttttccag catttcttag ttgagttagg 69120
gctttcatta ttacttattg ccttataatg cagtatgaac acttaccaca attcgttttt 69180

ctacccaacc tcttggcaat ctcaaggctt caaactaact ttgaaggaac tcttaattta 69240
tgtattacat ttcttattta tttatttact tatttatttt tgagacagag ttttgctctt 69300
gtcgcccacg ctggagtgca atggtgcagt cttagctcac tgcaacctcc gcctcctgcc 69360
tcctaagtag ttgggattac agggatgcgc caccatgccc agctaatttt gtgtttttag 69420
tggagatggg gtttcaccat gttggtcaga ctggtctcaa actcctgacc tcaggtgatc 69480
tgcctgccat ggcctctcaa agtaatggga ttacaggcat gagccactgc gcccagccaa 69540
tttcttattt ttaaataagt gagatggtac ctttttctc tctgtaacag aaaaaaaaaa 69600
agaaatatat gaagtcatac aatgtaacag aggtcaggag tttgtctttt ttttttttg 69660
agacagagtc tcactttgtc acccaggctt tagtgcagtg gtgcagtctt ggctcactgc 69720
aacctctgct tcctgggttc aagcaattct catgcctcag cctcctgagt agctgggact 69780
acaggtgtgt gccaccacgc ctggctaatt tttgtaattt tagtagagat gtggtttctc 69840
catgttggcc aggctggcct caaacccctg acctcaggtg atccgcctac ctctgccttc 69900
caaagtgctg ggattataga cgtgagccac tgtgcctagc caggagtatt ttttttatt 69960
gtggccacat gttatacctt tttttaaaaa aaagaaaaat ttatttcttt gttgtactat 70020
tttaaataaa tacaaaagta ggatttttaa agattttctg aaatgtatat attgctttca 70080
taatatcaac agaatttggt aagctttact aaaaagaaac aaaagcagct gaatggagag 70140
aaactggaaa acacctcacc ttggtaaaag agataagcta gaagagaggt gtaattttta 70200
tatagttgta agattgccat taaatatgca tcattcattc tttcagttcc ttggacacag 70260
catgaatttg agttggccca atgggccttc cggaaatgga agtctcatca gtttacttca 70320
ttacgggact tactctgggg caatgccgtg tacctaaagg aggccaatgc catcagtgtg 70380
gaactgaaaa agaaggtatg gagcaggagg acacaggaga gctggaggca aagccgagcc 70440
tgctgtgggt gcatctgggt tctcaccttg aattaacctt tcctttgggg gaactcagct 70500
gctttgtgct ataaaacagc tttatatatg tctttattta atatatgtgg ctggaacata 70560
gtaggtgcta attaaatttt tgttaaataa aatttttaaa acatgttttc tagcttattt 70620
tacaaataga ctatattcct acaaaaactg gagaaacata gtaatagaaa aatacttctc 70680
ttgggtttct gaataatttt aatttttcttc cacttttctg ttattaagtt tgcacctcta 70740
agtttggtat gccccctttgc cattgatctg cttctcctct agggtctgag aggcactgct 70800
gtgtagtgga caagcacatg caccctggac ccagacagtc caggttccag tcccagctcc 70860
acgagtggtg ttgatccccc tagccctgtg cctccattcc cccatctaca taatggggtt 70920
ggcagcagta cctggcacca agggattttg tgaggaatga aaatgggcat atgcataaag 70980
cacttatagt gtctggcatg gtaaaaggca ataaatgtta ttcatttttg taagtattac 71040
caaaacttca caccaaacct tagattcagt cttctgtatg cccccgtgtt tatcagtttc 71100
ttcactgcta agattacagt ggaagaccat attgtgtagt aggtaagagc atggaccacc 71160
taactttggg gaaattaatt atcttctcct agaactaaat tctagtttta gggaatacga 71220
tatgttcttc attcataatc ttggcagcct ttctcatttt ataattattt attagccact 71280
cctctcctca agtagactaa ataaaggcag taactgtctt tttttcttac tactgtgcat 71340
agcatttaat taatattaaa taattgttta ataattgttg gtgaaactat aacgtggtag 71400
ggagaaataa atgatgcaat gtcgtgaagt acttaccata tttattaaac acttttataa 71460
aaccatattt attaaacact tttttttatg agataggatc tggctctgtc gcccaggctg 71520
gagtacagtg gcgcaattat ggctcactac tgcctccacc ttctgggctg ggctcagttt 71580
gtagagacag ggtctcactg tgttgcctag gctggtctca aactcctggg ctcaagtgat 71640
ccttccgcct cggccttcta gtgtgctggg attacaggtg tgaacttccg tgcccagcca 71700
aataaatgct tttaagagca gtagttgcaa tcctctgaat ttaaaatggg aaaccttgct 71760
gtgttttcat cctctctcat tagtgaggga ctagatatgc attttgtggt cgacctttg 71820
ttctctggta taaattctat aaaagtcttc attcattcaa tcaaaatgta tttataaaga 71880
ttttatgccg ggtgtggtgg gttgtgcctg tgatctgagc cgccgaggtg ggtggatcat 71940
ttgagctcag gagttcaaga ccagcctggg caatgtggtt aaaccccatc tctacaaaaa 72000
atttttaaaa atcagctagg ctggtgatgc atgcctgtag tcccaggtac tcaggaggct 72060
gagatggtag gatcacctga gcctaggagg tggaggctac agtgctctgt gattgcatca 72120
ctgcacccag cctgggtgac agagtgtaac cttgtcccca ccgtgcctcc cctcaagaag 72180
attttcgtat gtgcaaaaac actgccaggc taggaaaaca aagtccttaa atagcttaat 72240
ttgaagttag ggttattccc tatgactgct aacaggtggc catatttttc agtagaaagt 72300
tagtcacaag atatttgttg agcacattcg acctctttcc cattgggtat gattgtatcg 72360
tctcatctct gaagcttgct tgctgaaaac catttgtcaa tggtttattc tttctattca 72420
ggtgcagttt cagtttgttc tgctgactga cacactgtac tccccttttgc ctcctgaatt 72480
acttcccact gagatggaaa aaactcatga ggacaggcct ttccctcgca cagtggtagc 72540
agtagaagtc caggatttga agaatggagc aacacactat tggtctttgg agaaactcaa 72600
gtatgaaaac attcataaag gctggttgtt ttatttagga aataacaatg acctgctcaa 72660
gtgagctccc tccagctctc ctctctctga ggacactgtt gcttacttcc ctgtactttc 72720
taaatgctgt tgttggccaa tggtattacc atccatctgt tttcttttaa gtgaataaat 72780
gcactcctca cacactctga catttgtaga tatttaagtt taccctccat agaccctctt 72840
ataaatttta aatattccag tttcatcata tttctgttac agatttggaa cttatttatt 72900
ttttctatcc tggacctcag gataatcaat aataaacatg gaagttataa tctggcttct 72960

tttacatgta gtataatgag gtatttcccc cgtctcagtg tatttgagat gtcacataca 73020
gttgacctac aatattcccc cttatgactt ttcacccggt agaggatagg tggccattct 73080
tgacttttga tggcttatct ctagaatcaa gacactttag atactgagtg ttggtttctg 73140
tatttttggg cttcattctt tttagtaaca ttaaacagtg ggagcaactc ctttttttt 73200
ttttgaagaa aatcttgaac ggaagatgct tccaaagtga aagtttaatg caatctcatt 73260
atattaaccc aatgtgcttt gtgtttacta aataggcaga ggctggattt gatgcgagag 73320
atgtatgata gggcagggga gatggcctcc agtgcccaag acgaaagcga aaccactgtg 73380
actggcagcg atcccttcta tgatcggttc cactggttca aacttgtggg gaggtatgtg 73440
atgattttgt tgatgtcttc ttttaaaata atgattagtt ttaagtgttt aaaacctgag 73500
cagataatat aaaaagttaa gaggaaaaaa aaaggtgtaa aaaggcctta tctataatct 73560
tatccactca caaccactgc gtcatttggt atatttccct tcagttatat tcagcaggca 73620
tttttgtttc gactctcaaa ctttcagcca aagtatttaa atcttccctg actttaaact 73680
tacaaagtgt tctgtaattg attaagataa atttagaagg aaaagaaaac ttacgcctgt 73740
atcctacaca gtttaaattc atagtacagc ttatcagaca caatctcttc gtgatctgat 73800
ttgtttttct ttttttaaa aagcaattca tcccacttta gtcattgcaa gggtaaatta 73860
attgcacctt ggtagtcccc ttcagcctca ttttttttt atacattatt tccagcactt 73920
aatgaaaaca catagtgcct ctatgagtcc atattgagtc aagtgttctt agaggttttc 73980
aaattttcaa agcagggaaa tttaactttg agcttatgat ttcagaattt gaattgcttt 74040
ttcaattgca gaaaaccctg tgatgttttt taaagaggct gggtaaaaag gagttgctga 74100
agtcatagta tgttgtggaa atgaagtaaa tcttttccac tctgctctga tagcaaagta 74160
gtttttcaaa gagatgaatt ggccaaacac atcttaccct agcaaagcac atgctgcttt 74220
gccatatgcc atttcatcca ttttggtttc tgctccaaaa tgacgtgaga atttcagtaa 74280
agactgaaag ccttcttgtt tcactttgat gttgagtggg tcctttgagg cctggatgct 74340
gatcttttat gtgaactgga gaatgtaaac gtgttatagc acagtattta cctaggatgg 74400
tctactcacg tgtttccttt gcttttctta ttctaaggat tagtccttac ataacttttt 74460
caacttgccc ttgttccctt tcacctgacc caaatgattt atgcttattt tatccttttg 74520
cttttccatt tgcttttatt gtgttgattc cccagtggat tctgtcctgt tagcacctat 74580
tgcttcctgt ttaaccaact attcctctct ccctggctgt gttaattggc gtcttacctg 74640
gtgtctagct cccccatttt ccacggctgt gtgaacgagc gccttgccga ccgcacaccc 74700
tcccccactt tttccacggc cgattccgac atcactgagc tggctgacga gcagcaagat 74760
gagatggagg attttgatga tgaggcattc gtggatgacg ccggctctga cgcagggacg 74820
gaggagggat cagatctctt cagtgacggg catgacccgt tttacgaccg atccccttgg 74880
ttcattttag tgggaaggtt ggtgaggtta ttgtgagaaa ggcgaaaagg gaccagctct 74940
tgctctgaag gcctccctgc ttgcacaatt ttggataacc ttgcattagc caattcaact 75000
catgaatgct ctttttcaag ttctccaata cttcaagctc ttagtcagtg ctggtctggc 75060
tgagatggtt ctcaggcatt gtgtgtgagg tcctcagtct actcagcaca atacgtagca 75120
atatatgaag actgcatgga agacaccttg tcttaaattg cccagtaatt ttgttctgaa 75180
gaagggaacc ccaaaggaga aagtaaagct aagagtcaaa cgtagtgggt tcagtgaata 75240
gtattatggt ctagatgttt ccttttttaaa ggattgtgat tgaaaaagca tatggaggta 75300
acatgagaga gggggcagag tgaggccaag gtttggctga agaatttgtg gtactgcagg 75360
agtcagagga caaaagtaga ttcaggagta gagttaaatt tatttttgc ttagtctact 75420
gaatcctaaa aagaatcagg aaagttaagc ttctccttta cctaaaacgg atgcctgtct 75480
ttggccagtt tcactgagcc tatctacgga tctttggcat ctgctaaaat tttaacctat 75540
agggtttata atgataaagc tcagaaaatc atctattgac cccatcagca gtgaatattg 75600
gccaggcacg gtggctcacc cctgtaatcc tagcactttg gctcacacct gtaatcccag 75660
cactttggga ggctgaggtg ggtggataac ctgaggtcag gagtttgaga ccaacgtgac 75720
ccatgaaacc cgtctctact aaaaatacaa aaattaggct gggtgtggtg gctcacgtct 75780
gtaatccagc agtttgggag gccaaggcgg ccgggtcacc tgaggtcagg agtttgagac 75840
cagcctgccc aacacggtga aaccccgtct ctactaaaaa tacaaaaaat tagccgggca 75900
tggtggcggg tgcctgtaat cccagctact tgggaggctg aggcaggaga atcacttgaa 75960
cccggaaggc agaggttgca gtgagacaag attgcacaac tgcactccag cctgggtgac 76020
aagagcgaaa ctccgtctca aaaaaaaaaa aaaaaaatga atatttgttt ggtgctatat 76080
gaaactgctt tgtgttaaga ctaggtcacc aaactgaggc attgtctctg gctctcagtt 76140
tgtgcatcca tctttcagtg atgcagaact ctctatgcac atgatcctaa tactggtgtc 76200
tttgttcatc tccagccttc atgctgaggt aatttaaatt acaagttcaa cagtagagca 76260
tgcattgaat taaccaaaaa gtatgtccag gtttaaaagt attaaattta aatttaatgg 76320
tacattagaa tcacatattg ttgaatcctt tataaataag ggaattatgt cagtgttaaa 76380
tggatgagta ccagttatct aaaaattaaa ggatagcgag tctcctgacc tagtagagac 76440
ataggttcta aaatttatct caccaggtgc agtggctcac acctgtaatc tcagcaactc 76500
tggaggctga ggccagaagt ttgagaccag agtgggcaac atagtgagac cctgtctctt 76560
tgaaaattta aaaaacaatt agctgggtgt tgtggtgcat gcctgtagtc ccagctcctt 76620
ggaaggctaa ggcaggagga taatttgagc ccaagagttt gaggctgcag tgagccatga 76680
ttgcaccact gcactccagc ctggggacag agcaagaccc caagtctacc aaaaaatata 76740

aaataaaata aaacagaatt tgtagtatgg tattggtcac atggtcggaa cctcataact 76800
aaattagaag atgatcattc ttatggaaag ggggaaaaga aaagcttttt tgatagaaat 76860
gaaagcctca agtgtaactg taggtcatag tattagacag tctctgggtt ttccagtggc 76920
aaaggagcat ttagaattca gtttttccat gaataaacat aaagggatta tgtggccatg 76980
aatattgttt tctactggca tatcgccatg aatagataac gtagaatttc agagaagctg 77040
atttttatta aaattgagaa aggtatgtac aaataatggg acaataatgc ttgcttcatt 77100
tgatggctgc ctgtgttctt cagcacgttt ccatcctgtt ctaagggagt tggaagttaa 77160
aacccagttt ctggccataa actttgtccc agtgctaatt acacagtttc taaaactgtt 77220
cgcttgtgtt ctataccatt gcagtgaaga gtaggcagct tcatcttacg taacatggtg 77280
aggagaagcc tgacctaatt gtcaggagaa gccttgggtc ctaggccatc tctgtcacta 77340
tgacaagcaa cttaaagtga agcctttggg ctctcaagtg taaagtgaat aacaggtaga 77400
aataaagagc cataaagatc tttttcaact ttcatattct gtcatctaca ctagcaaaaa 77460
tgaatatcct ggttaagact tctactaaga agcatttgta ttttgaggaa gttgtgtgga 77520
aactcaagac aatcccactt aatgtggaaa ttaactgagg atacatgtat tgttttctct 77580
ctagttttat agttttcttt cttttaatca tgtcagatgt aagtataggt ataaagcaat 77640
caactctatt ttgattttca cataaaatac ttatggagct gtggctcagg aattgatttt 77700
tccatttgtt taggttccca agttctaacc ttgcaagaaa ggactggggg aggaagtgaa 77760
ataatataat agcaatcagt aaagtgaagg atagctttcc atttttgccc ttgactacat 77820
ttctcttcta caaacaaaat ttacaataat gcaacatcac ccagttgtgt agtggtcatt 77880
gtatgtgtaa ggaaagtaac tactttcctg tttttgtgat tcacaagatt ggatttctcc 77940
tattggcata taattgccaa tataaatcac ttgtaaaatt tttagtttct ggctgggtgc 78000
ggtggctcac tcccgtaatc ccagcacttt gggaggccga ggtgggcaga tcatgaggtc 78060
aagagttcga aaccagcctg gccaacatgg tgaaacctcg tctctactaa gaatacaaat 78120
attagccagg catggtggtt tgtgcctgta atcccagcta ctcaggaggc tgaggcagga 78180
gaattgcttg aacccgggag ggggaggttg cagtgagccg agatcgcgcc actgcacttc 78240
agcctgggca acagagcaaa gactctgtct agggggaaaa aaaaattagt ttcttgtgaa 78300
attccattag taagtttgtt ttttagccat gactgcctct gctaaaactt tatattgcaa 78360
gtgaaaaaga gcttgtccct ttagctattc tcttatggca tgctgggatg ttatgaacca 78420
tccatcataa tttgggatga aatgtaaaat atttaagcta gtcttctggg ttaaaagggg 78480
gttaaagatg atgattctgt gatctctcta agtaccttaa ctgacattct ttctataagt 78540
agcaccataa ttttgctaat ttttcttatc tttatgttga ccagtttgtt ccacttttg 78600
aagtaacttc attgaacttt ctgctaacgt atcctttcca tattcccctc tctcatccct 78660
gttctccttc agtctgctag gtcaaagtat aaggattctc aaaaatgcag agtaattatg 78720
aaaattctgt ttaaaaagct tcagaataga atgtttatat agcattgatt tggagctaag 78780
ggctatattc tggtaacata acaaggtgct atctagtgaa tgtgtgattg tgtagtcatg 78840
gagctgatgt ccctgccata gaattacaga attatctaaa aagggaaatc tataataatg 78900
gccttcaatg ctcccacact tggactagcc attgccatag gcagaaaggc aatctcatct 78960
ttactgtgtg tgcaaacagt actttaatgt aatgtgtgct taatataagc tttctttaaa 79020
aaaaaaaaag gtggctcctg tttttgaata gctatttta agatagatat agttaggaat 79080
ctaaatgtgt tctatatagt taatatccat tatgaggtgg ctctgaaaaa tcaacctagt 79140
ataagttgga tggctttgct tttctgcttc tgttayaacc cattttcta gaaaagcttc 79200
cacttctact taaagataag aggcaaatca ttttcctgtt cccttatgca catagatgtt 79260
cacttaccaa atatttatag agtggcccag ccctctgcag acagtggtga gcaaaaccag 79320
acatggttcc tgccctcata gacttatagt ctgatggatg acacagacat aaattaaata 79380
attgcacaaa taaatgtaaa ttatagcagt gatgaaatag ttgtgtgtgt tgtcggagat 79440
cacataacgt tgggggacct gttatgagaa agtggtgttt gaaataagac ctgaagatgt 79500
gaaatcaaag gcgtgaggct gggtgtggta gctcacgcct gtaatcccag cactttggga 79560
ggctgaggtg ggtgaatcac ctgagatcag gagtttgaga ccagcctggc caacatggtg 79620
aaaccccacc tcaaaaaaaa aaaaaaagac atcaaaggca tgaggtgtag gcatcgagtg 79680
cagagaggac tccaggcagg ggacatgtgc gatggctctt taatggccca ggccttgaaa 79740
tgggcaagtc tagctggggt gcccataata gaggagggtg atctgtgtat gtgttgactt 79800
tgaaagtgaa attatccatt tgatacctta cgccaaattc tttgtctttt aaattctgtt 79860
gcagtagttg ttgtttgagg gtcttgttta gaacttttta gaacttttca ttcttacgtt 79920
tttttacacc agagcttcaa ctctctgttt cttagcattt ctctctctcc tgctttggag 79980
cagactgaaa gggaagtggc attatttcat ctgatktaat gttcatttgc agttatttt 80040
gttatatgtt tcatttgtat ttcatttatg ctctagttat gttaacattt gtttttctgt 80100
gatgtattat gtctgttttc ataattttca ttactttgag ttgaaatccc aaatacccag 80160
tctatcctcc tcctctcttt gttttaaaat agatttcaga gcaggaaaag catcagataa 80220
acatgtatgt cacttggttt tgtttgattt attgatacaa gtgtctgtct ttgtcttttg 80280
tcttatgttc gaagatcagc atataccaaa tgggaagact ggaatcatag taggaactag 80340
attctatttc cctttcacta attggaatga acactctagc tttgtgacca ccagttttta 80400
ttgctttttt aaaaaagtag tagctacaca ataaagaaga aaatacaaat gtgataggac 80460
tttgctagtt tgcttgaaat ataaagtgtt agagctataa tttagctttc acttatgcat 80520

ttctgaatta ggttttgaac tctaggagca aacgcaaacc ttttaggcat caactgtcat 80580
cttctccctt aacatttatc ttttttcttt ttttctcttg agacgaagtc ttgctcttgt 80640
cccccaggct ggagtgcaat ggtgcgatct cggctcactg caaccttcac ctcctgggtt 80700
caagcgattc tcctgcctcg gcgccccccga gtagctggga ttataggtgc ctgccaccgt 80760
gcctggctaa attttgtatt tttagtagag atggggtttc accatgttgg ccaggctggt 80820
ctcgaactcc tgacctcagg tgacccgcct gcctcggcct cccaaaatgc tgggattaca 80880
gacgtgagcc accatgcccg gcccttaaca tttatcttga tgcccattca gagtctcctc 80940
taatcttaga actcaaaaat ttcttgctgc agagcttctc tgcccatatg tgtctgctga 81000
cctcaccttg tcagtcaccc tgcctgaaga taatagtcat tttttttttt ttttttttt 81060
ttttttttt gagacggagt ctcactctgt cgcccaggct ggagtgcagt ggcgcgatct 81120
cggctcactg caagctccgc ctcccgggtt cacgccattc tcctgcctca gcctcccgag 81180
tagctgggac tacaggcgcc tgccactatg cctggctaat tttttgtatt ttttgtagag 81240
acggagtttc accgtgttag ccaggatggt cttgatctcc tgaccttgtg atcggcccgc 81300
ctcggcctcc caaagtgctg ggattacaag catgagccac cgcgcctggc ctctttttt 81360
ttttttgaga cagagtctca ccctgtcacc caggctagag tgcagtggtg caattttggc 81420
tcaccacaac ctccacctcc cgggttcaag cgattctcct gctacagcct cttgagtagc 81480
taggacaaca ggtgcatgcc accatgcccg gctgattttt atatttttag tagagacagt 81540
gtttcgccat gttggccagg ctggtctcga actcctggcc tcaggtgatc cgcctacctc 81600
agcctcccaa agtgctggga ttacgggcat gagccaccac tcccagccaa tatagttatt 81660
tcttaaagta acaataaatg gccaggcacg gtggctcacg cctgtaatcc cagcactttg 81720
ggaggccgag gcgggcagat cacctgaggt caggagtttg aaaccagcct gaccaacagg 81780
gagaaaccca gcctactcag gaggctgaga caggagaatc gcttgaaccc gggagataga 81840
gtttgcagtg agctgagatt gcgccattgc actccagcct gggcaacaag agtgaaactc 81900
catctcaaaa taaataaata aataaggtaa caataaagac ccataggcca ggcgcggtgg 81960
ctcacgcctg taatcccagc actttggaag gctgttgcgg gcggatcacg aggtcaggag 82020
atcaagacca tcctggctaa catggtgaaa ccctgtctct agtaaaaata caaaaaaatt 82080
agctgggcat ggtggcgggc gccgatagtc ccaggtactc gggaggctga ggcaggagag 82140
tggcgtgaac ctgggaggca gagcttgcag tgagccgaga ctgcaccact gcattccagc 82200
ctgggtgaca gagtgagact ccatctcaaa atcagacaaa aagcaaccaa taaagaccat 82260
atgctgtgat atggaattta ttttagagtt attaactgta tttcgtccca tttccctgcc 82320
ttttttgtct cctggtttct gttttatcta ttgggaaaaa tgaaacaaaa cctctttata 82380
atttcttaaa ttaaaaggtt aagctcgcct gggcacggtg gctcacacct gtaatcccag 82440
cactttggga ggccgaggcg ggtggatcac aaggtcagga gattgagacc agcctggcca 82500
acatggtgaa accccatctc tattaaaaat acaaaaatta gctggctatg gtggcacgtg 82560
cctgtaatcc cagctactcg ggaggctgag gcaggagaat cacttgaacc cgggagccgg 82620
aggttgcagt gagctgagat cgtgccactg tactacagcc tggcgacaga gcgagactct 82680
gtctcaaaaa aaaaaaaaaa aaaaaggttg agctcttgaa tgggctcctc ctccaacctg 82740
cccgagcagc attcatagag cagtatcttg aaagtctgct tccaggtgtg actcttctcc 82800
ctacagtagt taagttttag agcatgattt ggtgtgtgtt ggtcttcgtc aacagttctg 82860
ccaagcgttt gcttgttcat tatctgtgct gaaaaggcag ctggaggcaa atctgagagg 82920
ctaggactga gaaggggtct taagacaagg ccagaagtca gctcacagtt gcaggaagat 82980
gttctcagtg aatctgaaag ccagtttatc tctccatgga tgttctgaca tttgtctcct 83040
ggtttctgtc tttagggcat ttgtttacct gagcaatctg ctgtatcccg tgcccctgat 83100
ccacagggtg gccatcgtca gtgagaaagg tgaagtgcgg ggatttctgc gtgtggctgt 83160
acaggccatc gcaggtaggt gaccctcttc tgaaatgaga gctgtgagtt ctttgtctag 83220
agacaaagca ggctgattct gcgtgggtca cgcttatatt acacatacaa actgtgggga 83280
gtttgtatgt gtaatataca gacacatact ttggagaaag cctttattgc tcctttctag 83340
caattgttcc caagtatcac aaatagatcc tgcttagtcc agactactgc tactttgcaa 83400
taatggggta aagagatttc cttaaacatg ttaaagaaca gctcttgcat taactatata 83460
aaagtcattg gtgtgttgtt ttttttttttt ttgagacagg gtctcacttt gtcgtccagg 83520
ctggagtgca atggtgcaat catatctcac tgcatccttc acctccaggc tcagccatcc 83580
tcccacctca gccacctgaa tagctggaac tacagatgca caccaccatc ccagctaatt 83640
tttaagttta taaattttat agagttgggg tctcattatg ctgcccagca tagtctcgat 83700
cgaactcctg agttcaagtg atcttcctgc cctggcctcc caaaatgtta ggattacggg 83760
tgtgggccat tgcacccggc cagtcattgg ccttttattt tgttttgaga cggagtgtcg 83820
ctctattgcc caggctggag cgcagtggca caatctcaac tcactgcaat ctctgcttcc 83880
tgggttcaag cgattctcat gcctcagcct ccctagtaaa tgggattaca ggtgcccacc 83940
atcacaccca ggtaatttttt atatttttag tagaaacgaa gtttcaccat tttggccagg 84000
ctggtctgaa actcctgacc tcagatgatc tgcccacctc ggcctcccaa agtgctggat 84060
tacaggcgtc agccactgca cccagcctgg cctaaataat atatcttaac tgcttcatta 84120
ttaacaaagg attgttaggc ctttgtcttg cttagttgtt gttttttttgt ttgtttgttt 84180
ttttctcaga gtacttgtaa tagacatttg tttggctgtt gttttgtaat gcacgggtca 84240
agctgatata agctgattgt ggtggctcgt gcctgtagtc ccagctactt gggaggctga 84300

ggcaggattg cttgagccca ggagtttgag gccaacctgg gcaacatagt gagaccccat 84360
ctcttaaaaa acaattgaaa aaaaaaaaag attttggtat tttccatttt tccatgatcc 84420
agcctatgtt ctgaggaaag aatgtggtca tttgctgtaa tctttctgtc ttaattccca 84480
aatgctgggc tgtgtccagt gacctttcag cctttccccc atggaaactt gtgacacaat 84540
caccagtgtt gagtatatgg aagtctcctc cacaagtaca atataatatt ttcaaatact 84600
tgttgggtat tattgactat tatcttgcca gcatggagac agtcaggatt tctaggtatg 84660
acatttgaac aatctcaaga ataaaactga ttctttttt ttgagacgga gtctcgctct 84720
gttgcccagg ctggagtgca gtggtgcaat ctcggctcac tgcaacctcc gcctcccggg 84780
ttcaagcaat tctcctgcct cagcctcctg agtagctggg attacaggtg cccgccacca 84840
cgcccggcta atttttgtat ttttagtaga gacggggttt caccatgttg gtcaggctgg 84900
tctcgaactc ctgacctcgt gatctgccca cctcagcccc tcaaagtgct gggatgacag 84960
gcgtgagcca ccgcacccgg cctgaaaccg attctttatt gactagtctc atttgagctg 85020
aggaagttga taaatctaca tgaatagagt gcgaagcagc ccatgccgtg ttactacttt 85080
ggtatcattc tctcatgcca aagacttgag tttataatcc agcaagagct ttttccctcc 85140
ctcccccctgt gtagtctcac tcaattcttg ctaatttttt tttctgcttt agcggatgaa 85200
gaagctcctg attatggctc tggaattcga cagtcaggaa cagctaaaat atcttttgat 85260
aatgaatact ttaatcaggt gagaaaccgt caggaagaag gaaaaccctg tggaaagaga 85320
aagaatgttc agcattggag atcagttctc ctgtaaaact gaaaaataca gggtgctggt 85380
tgccacagag tagttcaata gaatatggga catgtctaac aattatgaag gtaaaggtgt 85440
atagactcaa tattcatatt ttattcattg cctttcctca gacagccctc agtccatata 85500
attttccagt cactattatt tgttggacag ataaacagaa gtaaggcaac tggggaaaaa 85560
tacgttttta tactacttta ccatgtatct gcccctgcct ttttttcaga gtgacttttc 85620
gtctgttgca atgactcgtt ctggtctgtc cttggaggag ttgaggattg tggaaggaca 85680
gggtcagagt tctgaggtca tcactcctcc agaagaaatc agtcgaatta atgacttggg 85740
tatgtagaca tagtttactg tgcttgggga cattttcgac aaagggaaaa tattgaccat 85800
tatcaaggga catagtggcc ttcatcaact aggaatggaa agcatgccca actccctcct 85860
ctttgcatta tttgaaccat ctgtgtcttc atttgaccct ctttagattt gaagtcaagc 85920
actttgctgg atggtaagat ggtaatggaa gggttttctg aagagattgg caaccacctg 85980
aaactgggca gtgccttcac tttccgagta acagtgttgc aggccagtgg aatcctccca 86040
gagtatgcag atatcttctg tcagttcaag taagctgccc ctttgctctg cctcccagct 86100
agctgctaac cgaggtgaca tctcttgtgc actgtagagt gctttacatg tgtgagggat 86160
taacactctt gagacaaaga ctgatcagtc tctgaaggaa aatactttca tcactcctat 86220
gggagtgaga accagaaacc tgtcctggaa tgcaacaggg ttccctggca ggaggagcat 86280
gctgcacatc ctgctgcagg gctctgctgt gagcggtccc cctttaccag gaatgtaggg 86340
tctgatcaca taagcctgtt gatgagtctt tctagtcctc ttcatgttga aatgtcttat 86400
taataagcct gaaagcaatg acttatctag ctagcaatta cgcctattta atactcggga 86460
aaaggaaaat ttaaacaatg taaagaagga taaattctac atcaactgaa aaaagctagt 86520
atgtagggaa gatgtgatac ctctcatttc tcatctacat aggacttgct ctttcccgcg 86580
gccctctcaa accctgtgct ctggatgctt tcttctcaag actgccccag gaaaacattt 86640
ttttgtgaat cttcagtttt aatacctcca aacttcaatt tgtcttttca gtacttggaa 86700
agggccacag tacaagagac tgaagtgtga cccctccata cctaccgccc tcagtttgcc 86760
actggtgcca ccaagaggct gctttgggtt ctgccatccc tctcccctca gaatgtgcct 86820
tggccgtccc atgggtccta gtgcagttta taaagctttt gtataggttt tatcacagaa 86880
aattgagtca aatgtgattt aaattaaatg tgatcactat atagtaagcc actaatatca 86940
tttgctacaa gtcatagttg ttggaatatt acccattgag cattgatttc acctatttaa 87000
tgtgttttca ctcataccct caaacttagg gcaggatagt gttcaccaga tgaatctgtt 87060
ttttggaact gttgctatta ttctaatcag gcaactgata cgttttatgg aaaacatcta 87120
tttggttaac tgatgtctga agtgacaata atagtgagtt gacctttgtc ctacaatctc 87180
ctaaaaagag cagcatctct acctggtgtt agctctgggg aaaagggcat gttgcttctg 87240
ttgaaagcag cctcaaaggc gtcctgtgtt tttatagtcg ctcctctggc ctgtgccggg 87300
aaggtatggc tgcagagcac tgtgaatgag ttcacctgtc gtcagggaac cccaggaatg 87360
ctgcggtgcc agcaatacag gaatcaggga tttggagcac ttggcaaaag catgctagtg 87420
ttctcagacg ggtgaactaa tttggaagac aacaagttca taytaccagt ctgtctgtcc 87480
ccccagttga aaatagtttc tttttctcca ttttttcttg atggaagggg tgtccttctg 87540
atggattttt ctatctggca gtaaatcttt gtgttgggtc ctctttaact gtgcagttct 87600
ttttatttca gactaataat agggagagag agagagagaa cgataacatc tcactrtctg 87660
tgaaaataac taaatatagc ccattgtttt aacaaatctt tccagagtat cttctgtgtg 87720
tcgagacctg tgggaaaacg cttgctgtct gggggagaga ggcatataaa tacattatta 87780
cagtgaaatc attgctatta aggaaaaaaa cattactagt aatgtgggaa cacaaagctg 87840
ccagggtaag tcaggaaaga ctgcctggaa gaggtgacat ttgaatccac aacagactct 87900
tatgtaacta tgccatccac aggctctcgc actaactgtg gaagcttgag tctgctaaat 87960
tttgtctgta caaaatgaca ttgtttttat ttttagtaaa aagctttac aaaggaagag 88020
gagttaattg aatgtacttc ttcactctct ctaaagcatt cttgcatata cctccatata 88080

gctaagtgat cagggtcagt taatttttca gtcatttttg aaataaaatc taaaatgccc 88140
tctgtgcaat atggaaaagg ttcttgacaa ggaaagtaaa taacttaaag aatctgctta 88200
agtggaatgc ttctttctgt aggatccaca ttgggctctt gaaagagatt ctgataaaac 88260
cgtttaatgc attgttcacg agtcttttta ttttttttat gaagcttttt gcatcgccat 88320
gatgaagcat tctccacgga gcccctcaaa aacaatggca gaggaagtcc cctggccttt 88380
tatcatgtgc agaatgtaag tgacatggac ctttttgcca aacatatgtt tttctggtac 88440
cttggctttc tcagcccggg aagggtgata agaagaggta caagtcactg gctgtgcagg 88500
gggagagtag acaatagagg gcgtggctag agtggtgagg tcatccacag cggcctgagc 88560
agcaagcctt cttgtctgtg ttacagtggg cctagtgaat gtgggttcgt ctgccaataa 88620
aatcttccat cagtttgaat gcaatgagtt gttttagtgt ttctgtccta gaaaaataag 88680
catgcttaat aagagtttta ttaaagtgaa tgcattgtcc tttaaaccct aagatactta 88740
tgttgctcga ttgcaaataa cataaaaaat atcaaatacc ttctgtagct aaattaagat 88800
ttagatctgt tttccttgca actgatagta taattaacac attaatgccc tttacaactg 88860
tattcctacg tggagtaaga aaacagaccc ttccctctct tacgtgaaat ttcattcaaa 88920
ctctattcca aagtcattat cctcatcaca atagctaata ttcactgaga gtttatttg 88980
tgttagtcac tgttctgaga gtttttgtgt aatttaatct ttacctttat accatagatt 89040
aagataatcc ctgtgttagc tgagatgact gaagcataga gcaattaaag tcactcatca 89100
ggctgggcgc agtggctcat acctgtcccg gcactctggg aggccgaggc aggtggatca 89160
cttgaggtca ggagtttgtg accagcctgg ccaacatggt gaaatcccat ctgagacacg 89220
agaattgctt gaacccagga tgggggggtt gcagtgagtt gagattgtgc cactgcactc 89280
cagcctgggc aagaaagcga gactcccatc tcaaaaaaaa taaaaaagta ataataaaaa 89340
taaagttact catcagtgtt acaagaggca gagataggac ttaagctcag agaaatttag 89400
ctctggtatt caggatattc actactacat gaaactgtct ctcattctct tgctaaccat 89460
gtctgctctc taaagtaaga gaggaaagga gaacttcaaa gatggaacaa gcaaacacaa 89520
aatatgccat cttctttttc atgttggtat aaatataagc aagcatgtat ctgataaact 89580
ttgatcagct gctgtgcaca aggattgtgc ctgacaactg gggaatgcag atgaatctca 89640
gatgctccca gtgagcttcg aattacgata ggggaaataa aatgcataca taaacaccag 89700
tcatgcaaag atgaacatga gaaaatattc actaaatgca gtgggagtgg agaaagggaa 89760
agttatttct attaatggaa agctatttta agttactttt tgttggggag gatagttggg 89820
tggagaaagg ggataggaag ctccctgaag gatgtagtat aagcaagaaa aggtgtgtca 89880
aacttggcat tgtgcctggc acataggcca tagtcagtac atgcatggat gtgctggaag 89940
ggccttccag gcagggttgc agcctgggcc agtccttggc acagctgagg agagagcagg 90000
tgaaaggagc agcaggtgag tgagtgagcc agaatagtaa gtgtgggaag gacaacaggg 90060
gccaagatca gttgaggcaa agtatggaag atcttgaata cctgaagctg ggaaaagtta 90120
ctttagaatt aaaaacaaca taggggatcc cctgtttttt gttgcaagag tcctgcatag 90180
taaaataccc agcagatact ggacactcag cagcagtaaa attgggccag gcatagtggc 90240
tcatgcctgt aatccagcac tttgggagct gaggcagaag gatcagttca gcacatgagt 90300
tcaagaccag tctgggaaac atagtgagat actgtctcta caaaaaattg ttttaaatta 90360
gccaggcatg gtggtgcatg tctgtagtcc tagctactcc ggaggctgag gcaggaggat 90420
tgcttgagcc cgggagggtg aggccgcatg agccttgttt gcgtgctaca gagcaagacc 90480
ttgcctcaaa aaaaaaaaaa aaaacccaaa atacgtacta aagttctgat tgaagcaaaa 90540
atgtgtattt tcttgtaatc ttttcctaat cttgcttggc tttagattgc agtggagatc 90600
actgaatcat ttgtggatta catcaaaacc aagcctattg tatttgaagt ctttgggcat 90660
tatcagcagc acccacttca tctgcaagga caggagctta acaggtttgg accagataag 90720
caaacatttt tgatggtatt gttttgattt ttagtttctc aattgctgga aaggaagaaa 90780
tcttttgagg gattaagata actaaataag tattctatat taccctgttt cctctttaga 90840
aacttcacta tcatcttata gccacctaat cactgatttt agatagagca aattaaattt 90900
tgaccaggtg acctcaaata ccatcttttt gtaggaagag actttgaata tggtagaaat 90960
ttgaggtaat aggaatataa catgaaaatt gttaaaaatt gttcctaatg cccacagctt 91020
tatgaaacag aagctctaag tacagtcatc acaactgctg aggaaagcaa cagtgggcct 91080
ttctgatgat ggcagagtgt tggtgttggc agaagaattt ttttatttat tagctaagac 91140
ttagagggag acatttttta aatgagattt agcctgtttc actatctagg attatctcta 91200
ttatctggta ttatgccaga gttgaaaatg gaaaagtctg cttccttaaa gggaattaaa 91260
acatgctatt gataccaaca tccaactctt tgtagtcttt ttctttctct tgaaaaaaca 91320
gggtttattg atttctacac tgcaaatagt cattacaaaa agtttttttt cttttaaata 91380
aattcacaca aagagaatta gaaagcgatg gtaatgacca gcaagaggaa taataattac 91440
gttcatctta atgtctgtgt gccagttctg tttacgttaa tgttggaaaa ctccaacttg 91500
gaatccagaa cttcaaatct gcaaatagaa tgtcttgaga taggcacgtg gaagtcaaac 91560
agtttctctc tttttccctg ctttcagaag ctatacataa aatgttgttg ccttctgtac 91620
tgtcacagaa cttttacgta cattctcact cctacttgtg aaacccctaa agagaaggaa 91680
attcaatttg cagtccagca caaaggggag aatttctaaa ataaataact ctttatagtc 91740
ttgatctttg tcttcctttc tttgcagtcc gcctcagccg tgccgccgat tcttccctcc 91800
acccatgcca ctgtccaagc caggtgagca ctcgctccgc tttttgcatg atgatctctt 91860

tgtgaataca tgtcttattc tgaatgactt ttcaaataga agtcatgttt gaattcctat 91920
tcttctatat atccacatgt attcctgctc ctcttgtatg tatctagtag gaactagaat 91980
tcctagttcc tgatttaaaa taagaagtaa aggctgggcg cggtggctca tgcctgtaat 92040
cccagcactt tgggaggccg aggcaggtgg atcacctgag gtcaggagtt caaaaacagc 92100
cagaccaaca tggtgaaacc ctcgtctcta ctaaaaatag aaaattagcc gggcatggtg 92160
gcacacgcct gtaatcccag ctacttggga ggctgaggta ggagaatcgc ttgaaccggg 92220
gagatggagg ttgcagtgag ctgagattgc acctttgcac tccagcctgg gcaacaagag 92280
cgaaactccg tctcaaaaca atataaaata aaataaaata aagtaataaa ataagacaag 92340
aagtaagaaa caataggggg ccttatctac aaactctctg gtcttctatt tgcagatttt 92400
gataatcact ctgttcctaa atctttctgc tttgcatgta acttcccata actgtcaggg 92460
gactaggcat ttctttgcat agtaaattta ttctgtatgt tcagtgatct tgttttgatt 92520
tttctttatt tctgtccttt ttcctcattg cctattcgcc aaattcccac cccaccctgt 92580
gctcactctc caccctctgt attccccatc ttcatctgcc ttcagaccat gagagaaaga 92640
ttgagctgat tcggtttctc gtgtctggct atgtgaatgt gcatatgttg gacacatttt 92700
ctgagcacgc cagtgtgctt gcatcctccg ctattgccac cttccaggat gaggctgacc 92760
cattgccgcc ttttctcttt ctgccagttc ccagctgcca gagcgagagg gtacccaaac 92820
gagatggtta gtagctgaat tagctccacc ttagatgtag aaccatcctc aggcgcttgt 92880
cctgaagcct ggctctctgc atatgaaaca ctgacatttc cctttgcaac ttgagctgaa 92940
tgtagagctc agcacatttt caaggttgta tcacattttc ttcctagaga tttgggaagc 93000
agcttctcta catgtgcctc tttcacatca ccatcgcagt gggggaggcc tttgcatttg 93060
tttttatgga ttctttgcaa tagtattttt aattttgaag attttactag tagttggttt 93120
gttagaatac ttctggtttt gctttctgac ccttagatga ctgccaggat cacttaccct 93180
attttctttc tgactagcat caccaactca tgtgaaactc agctgggctg aatactgctg 93240
atagggtaca cgtgtgaata gtgctaaatg attcataaag ctcctttatg tacattatgt 93300
tagtggatct tcatgattgc ttttcaagat agtaggctaa atattagggt ctacagttta 93360
taaatgagga gacagactcc cagaggttaa gcagcttaac taaagtcaca tagcatcagt 93420
acagttaaga atatggagtt taaatctggc cgggtatggt ggctcgcacc tgtaatccca 93480
gcatttggga ggctgagtcg gacagattgc ttgagctcag gagttagaga ccagcctggg 93540
caacatggcg aaaccctgtc tctacaaaaa aatgcaaaaa ttagccgggc gcggtggtgc 93600
acgcctttgg tctcagctac tcaggaagct gaggtgggag gatggtttga gcccagaggg 93660
cagaggttgc agtgagccga gatcacgcca taaaaaagaa aaaaaagcat atggaattta 93720
aatctgaaga aaatgtgtga cataagatca tttacagaat cctgtattta taaatgatac 93780
cagtagttgg tttccaactt tgtcctcaca ttggtatcac ctgggaaact ttacccctca 93840
tggaggaaga actttgtctt tcccttagag attgtgattt aattgatctg ggatgtgtcc 93900
tggactttaa ggtccccagg tgctcctatg tgcagggagg tttgggaacc actggggtgg 93960
aggatgggtc ttttcttcag gacatcagct ttgattgagc ccctgtgacg tgaggctttc 94020
caagcattgt ctctcccacg ttagtcccat gctttacgtg tcacacaatc tgagattctg 94080
gaattggaag gtttcatgag cacatttatt ctacacaaaa gctgttttct catcattaga 94140
gacagtttgt tgagaaaaaa gatactgatt tctttatttt ccttgttctc tgaatgtaga 94200
gataactaat aacttacaaa gtttatattg gtgtttctct gggtcttttg cattgaaatc 94260
tttcctctga ctcttgctgc ctttcccaag tatgtcttta aagactgaca ttagagaagc 94320
ttgtattttt gctgagtagt cttggaccag attttgacat actctaaaaa cttttaaaag 94380
ttccagccac caagttaaac acgatgagca aaaccagcct tggccagagc atgagcaagt 94440
atgacctcct ggtttggttt gagatcagtg aactggagcc tacaggagag taagtccaac 94500
ttaataaatt tttaaataag gcaaaatgtt tcaaattagg aaatgcaatt ttgaatcttc 94560
ttgtatttgg gttcatgaat cattgtcact gtcaaattcc agggatggtt ttgtttttgt 94620
ttttgtttta atagagatgc agtcttgtta tgttgaccag gttggtcttg accactcctg 94680
gcctcaagca gtcctcctgc ttcgacctcc caaagtgctg agattacagg catgagccac 94740
cacacctggc cgaattccag ttttgaattg atttttggtg gtcttggtat ttgaatcaga 94800
tgtttatcag agaactgact gttgagcata ataatttgcc ttctcgttag caacagcaaa 94860
atcatggaga caggcatctt gttgtttttt ggggtttttt ttgttttttg ttttttttga 94920
gatggaggct tgctctgttg cccaggctgg agtgcaatgg cgcgatctca gctcactgca 94980
acctctgcct cccaggttca agcgattctc ctgcctcagc ctcctgtgta gctgggatta 95040
caggtgcgtg ccaccacgcc aggctaattt ttgtattttt agtagaggca gggtttcacc 95100
atgttggcca ggctggtctc gaactcccga cctcaggtga tccacccgcc tcagcctccc 95160
aaagtgctgg gattataggc gtgagccact gtgcccagca gcaccttgtt gtaagggttt 95220
tatcataatg ggttttatta gtagtaaagc aaaaggagca tattgcatag aatatggtaa 95280
agacttacat tactgatgtc aaggaggaca tagaagggca gaatattttc ctagaccagg 95340
gtttctcaac ctcagcacta tttggggctg gataattctc tgttgtgagg ggctgtcctg 95400
tgcatttaga atattttgca gcatccttgg cccctaccca caagatgcca gtagcagctc 95460
cactctattt gtgacaacca aaaatgtctt cagacattgc caaatatccc ccagggaggg 95520
catcctcagt tgagaaccac tgtcctagat caaagtgata ccatcagctt actctgtctc 95580
gatgagctga gttagagtaa gtgattaatg taattgttgt gtataatcat ctctaaaatg 95640

cttttgtaaa aacacaaatg aaagagatct tactcttagg acagaggcaa ggggctgtat 95700
cagttaatat aaagggtagg agtaaatgtt atcctaatgt tctcctatat gtaattctga 95760
gaaccaccat catggtcatc attgggaaac acttattaag ttcctactat gtgtaggata 95820
ctgtgttaag tgtggtatgt agttgttagt agtgacagtg ctttgaggaa tttgactttt 95880
catagacaga cttatgatgt gattacaagt tgcattatga gggagaggta gttacagtaa 95940
ttgtttctac tgaattaatt gctttctgaa tcttaaaaaa aactatagtc tgaaattatc 96000
tgtgtttctt atctcctgtc aattgccatt gcacttctac attatttatt gaatacttaa 96060
tatcaagggg ctggatgcaa ccaaaagatg ttactgtcct ttgactactc aagtctagag 96120
aggattgaga catagatgtg tggaaaggat gagtaaatgc aaatgactgt atgattaaat 96180
gccagcatgg gtgagaaaga catgggatca gaatttatca aaacaaagac caccaggggc 96240
taagccttgc agatgaagta gcacttagtt ttttcttttg cgtttctatt ttttaggtat 96300
atcccagctg tggttgacca cacagcaggc ttgccttgcc aggggacatt tttgcttcat 96360
caggtactaa tgagggacca aaacaggcat cggagggaac actgaaaaga atttagaata 96420
agagacggaa ttctttctta ttctctgttt tcatctctat ttcagagggt gggcggggca 96480
ttgtcctgtt gtcatttttt tttttttctg tgtctgtctg tagatatttc ttcctcagac 96540
cttggcttat atttatcctt gcttttatat gtatttagaa tagctatttt gaaaaatatt 96600
ttgtttcatt caattactct aatttgtgta gctcctatgt cacttttaat acttttcatt 96660
tcgtgatgtt tggaagctct gtgctttcac accagaagtt cttaatagga atagaggatg 96720
aaggtttcaa tggactaaaa atctggttat gaaacatagg tccgcttttg gtaaatcaaa 96780
agtaggttca gtcagtgttc aaaagtactt aagacccaag tgggcatagg ggctagggat 96840
accatacatt tgggacatac aatcagggca tggttgaata cttattatac aaggggctgg 96900
gatgcagcca cataatctcc tgtgacagat tatggttatt cttaaaacca tagagaaaaa 96960
gaatttttta aaaaagaaaa gaaaaccata tcagaacctg actgttcctt agtccctgac 97020
acatctccct ttgttttgaa taacaatgtg gatagcattt tccgtctttc cacacctctc 97080
cttcctgctc atcccccatg ccatgccata caggccagaa aacagtagac gagaggagat 97140
aatagattgc ttcagctaaa ttgcaaccct gcttcattac ctagggcatc cagcgaagga 97200
tcacagtgac cattatccat gagaagggga gcgagctcca ttggaaagat gttcgtgaac 97260
tggtggtagg tgagtacgtt tcatcagcca aggatagaac caggacttac agagattttt 97320
tttttttttg agatagggtc ttgctctgtc acccaggctg gagtacagtg atgagatcat 97380
agctccctgt agcttcaaat ctcctgggct caagcaatcc tcctgcctca ggttcccgag 97440
tagctgggac cacaagtgtg aggcactgtg ccctgttaga gatgttattc ttccctgttt 97500
gcggtacatg gaatgtgtag aagaccatat tcggaaacac agactatgga gatgcagagg 97560
ccttgagttt gaaactctgt attcctgttt cttcgctctg tcattagtga tatgttatcc 97620
aactaactaa ctcagttacc agggaatggt tctgtggtag agtctgtgag agggcataaa 97680
agaaatctct gctggctggg tgtggtggct catgcctgta atcccagcac ttttggaggc 97740
tgagatgggt ggattacttg agcccagggg tttgtgcctg gccaaggtag caaaatccca 97800
tctctacaaa aaaatatgaa aattagctgg gcatggtggc acgcacctgt agtcttggct 97860
actcgggagg ctgaggtggg aggatcgctt gaacttggga ggtggaggtt gcagtgagct 97920
gagatggcac cactgcattc aagcctggac aacagagtgg gactccgtct caaaaaaaaa 97980
aagaaaagaa aaatctctgc taattgtgag cttgcatttt agttagggag acagcatata 98040
tatgtgcaga atgctttgat gaataaatac caaaatgagg acaaaggaaa atctgtgcta 98100
agagagttaa cagtgaaaac aagccaggag gactgaatag attcaagaaa aattagaggg 98160
ttttttttt ttttccagac acaagcattt tgagtaagag agtcaggaat atgtctaaca 98220
tgcaaatgac aagaaacccc atggccagag tggagaaagt aaattaagga gtattggcta 98280
gatagtgtaa actttagaat gacagtccga ggcaagagtt tggaactttt gttgccatgg 98340
acctctttgg cagtaacgaa gccttagaat tatgttttca aatgtgagaa ataaagtatt 98400
taggattaca aaggaaaccg ataatattga aatacagttg ttaaaacatt aagctcttct 98460
ttactgatgc atgaaataac aagatttagc agttattcta attaccataa tttccaagta 98520
gtgatcaatg aaaaagatat cttgtgtcca ggcacggtgg ctcacgcctg taatcccagc 98580
actttaggag gctgagatgg gcggatcaca aggtcaagag attgagacca tcctggccaa 98640
catggtgaaa ccccgtctct actaaaaata caaaaattag ctggttgtgg tggcgggcgc 98700
ctgtagtccc aggtactcgg aaggctgaga caggagaatc atttgaacct gggagacaga 98760
ggttgcagtg agccaagaca gcagagtgag actctgttta aaaataaata aataggccgg 98820
gcgtggtggc tcatgcctgt aatcccagca ctttgggagg ccaagatggg cggatcacga 98880
ggtcaggaga tcgagaccat cctggctaac acggtgaaac cccgtctcta ctaaaaatac 98940
aaaaaaatta gccaggcgtg gtggtgggcg cctgtagtcc caactgctcg ggaggctgag 99000
gcaggagaat ggcatgaacc tggaggcgga gcttgcagtg agccacgatt gtgccactgc 99060
actccagcct gggcaacaga gcaatactcc atctcaaaaa aatatataaa taaataaata 99120
agatattttg cacatggaaa catctgtggt ttctactggg gacagagtca cgggtactgg 99180
ccaattcttc tgtggtttat tgccttcctt tgtgattgaa ggagatgcca atttttagtt 99240
aaaggtttgt gaaaataaag atgtaatttt tttccccatc caagctcaag gaccctctga 99300
attctgtcca cagagtgctt gtgtctctct ggaccagatt aagaaccttg gggccaggca 99360
cagtggctca cgcctgtaat cccagcactt tgggaggcca aggcaggcgg atcacttaag 99420

ttcaggagtt tgagactatc ctagacaaca tggcaaaacc ctgtctctct acaaataagc 99480
tgggtgtggt ggctcatgcc tatagtctca actactcagg aggctgaggc tggaagcttg 99540
cttgaacctg agaagtcgag gctgcagtga accatgatca tgccactgca tgacagaaca 99600
agaccctgtc aaaaaaaaaa aaaaaaagaa tcgttggaag ggagattatg ttctctcttg 99660
tctggggcag cagtggtagt agcttttcaa tatttttcat aataaatttt tctttttta 99720
atttagtttt ttttttaact ttatttacaa agtactgaat gtcctttgac aaaataatgt 99780
ttatatggaa ccctaatacc gaaacagaag agatgctttc tgaccagaga aggagaggga 99840
ggatgcagag gcagcagtta ctatggcttc tctgaaggat cctatgatat atttttagtt 99900
tgttttatta acaaactctg atgtagtatt taccacatgc caggtgcttt tctgaatgct 99960
ccaaaaatac tagtttgttt aatcctcata gcaaccatat atggaaggaa ttgttattct 100020
cgcttgacat atgagggaac tgaggctcag agaggttgag ccttaggctc aaagcctcac 100080
aggtggtagg agttggaacc caggcagtcc agcactggag gccaagtcct taacgtctat 100140
gcaatgctgc atccgctagt gttaacttga aaacagagtg taggattttg agcaagaaag 100200
gggtgtaata aagtcagcat ggccaggtat ggtggctcat gcctgtaatc ccagcacttt 100260
gggaggccaa ggcaggagga tcgtttgagc tcaggagttc gagaccagcc tgagcgacat 100320
agtggggccc cgtctctaca aaaaaataaa ataaaataaa atagctggat gtggtgctgc 100380
aggcctaaag tcccagttac tcaggtggct taggtgggag gatcacttca gcccaggaat 100440
tcaaggctgc agtgagctat gatcaaacca ctgctctcca gcctgggaga cagaccaaga 100500
cattgtttca aaaaaaaaaa cacaaaaaaa caagaaactg tcagtcagtg gcttcagaag 100560
ttgctgttag gtgagagtgg aatggattgg aacggggagg tgtgaatagc aggtcagaag 100620
ttgctgttag gtgagagtgg aatggattgg aacaggggag gtgtgaatag caggtcagaa 100680
gttgctgtta ggtgagagtg gaatggattg gaacggggga ggtgtgaata gcaggtcaga 100740
agttgctgtt aggtgagagt ggaatggatt ggaacgggga ggtgtgaata gcaggtcaga 100800
agttgctgtt aggtgagagt ggaatggatt ggaacggggg aggtgtgaat agcagggaga 100860
ccagttaaat ggctgttgca ggttccagac cagagaggag gtggcctttt gtgatcaggg 100920
caatgatatt gtgaataaaa gttgggatga atgtgggaaa ctctttgaat aaagaatcaa 100980
tatggtttca tagttaactc aatgtgggaa gtaatggagt cagaccgtgg aggaccctac 101040
aggtaaccaa atgacccttc cagctcctga acttacacag agatttaaaa gtctcttttg 101100
ggcttaatga attcatgttt tgttttcaaa tgtgtccgtg ctctgttttt tttatccttt 101160
cttttaggtc gtattcggaa taagcctgag gtggatgaag ctgcagttga tgccatcctc 101220
tccctaaata ttatttctgc caagtacctg aagtcttccc acaactctag caggtgggac 101280
acccagagca gtgtgaagaa gtccacactt gcaggcgtta attggtacac cgttaggtgc 101340
cttattcatt aatgactccc agttcggaca aagaaattaa ctccccttctc ccttctagct 101400
tcaaaaatct ctttatttct tcacctgcct gctgtactct ccaaaaagaa agaaagaaag 101460
tattgcagat atttgtatgt gatcagttac tcttagagaa tggaagtgat cctgtcccat 101520
gtgaagtttg aatagatgta acaagtgata aatgaaaatt ggagaaagaa aacagtatat 101580
tccccaagga tttaaaagta craattaatt attgccaaga ttaaattttt tcctgtgaaa 101640
tggttgctgt ggagagaaag gtttctctca ctccccaagt atcatggaaa tgtgccctct 101700
gagataaaat gaagcctctt gtgttaagtc tttgctgctg ctgactatag gctccattat 101760
ccttagtatt tgtttttcat ttatgcacaa aagcgataaa taatatgaac tcttatgacc 101820
aagattggct gccataaata atagatttta ctttgttttt tattttaaat gtttcactta 101880
aaattatctt ttataatcag gtgataacag acagttcgta aagtacatag ccattctgct 101940
ttcctttgta agactagaac taaaccagct gggatcctct tacctcaccc aggaagggag 102000
ttctcattgt ctcagaggct ccgtctctat atcctagggg ctgggaaaca ggagcggctg 102060
cctcttctaa tgtgggctga gcacactgcc ttcccaccac aggactttgc tgtgcttcag 102120
tctgctcctg aaaaacttgt ttacatccat gttgagagga gggagggaag gaataagacg 102180
ataggcagaa atggaatgaa aggtttggac caggattaat ttctttctgt tgttttcttg 102240
tattatagtt taccatagga aagagaaaaa agaaacaaan ttctgtttgc tggcaggggc 102300
tgtggagaaa gagggaagga gtgagagggt acatatgtat gtctggggag ggggagtggc 102360
agaaataaag aggggtgctt tgactatcca tgggcaaaat cacgagcaat actatttccc 102420
tgaagtcagc agtaattctg catagctttc tgttttctgc caatgttcag ccgggtttta 102480
ccaccacatt ttctttgtag ctatgaaaaa tgtattgtaa atttccctag ggtacttagt 102540
aaaatatata gggaaagatg ggctttggtt gaaatactgg ttattctaat tttgcctcca 102600
ttagagtaac tttaagaagt cctttaattc acgcccaaca acaggtcagg tgctacaaga 102660
aatgaatttg gctttgttag aagcagctta tcttgtgagt tagaatgagt tcatcaaaaa 102720
aggctgaagt caagaataac aacaacaaca aaaaacacct taaatatttc tgttctcatt 102780
taaactatat aaatctacat taatttatca aacttttgct ggaaagccat ggccttttct 102840
ttaattataa gatcatttac tggagttctg attgtctttc ttgcaaaacc acacccataa 102900
ttcatcatct atgactttca gtttggtgga gcaagaatta gaaacctgtg aagaaaatct 102960
aagttcaaaa tcctttaaat tttttatcat ttattaacat taatcattta aacattttat 103020
cattttcccc tttcttttac atttgtctca cccaaactta acagttattt ttaaacaact 103080
caccattttc taacctttttg aaaacattca acttagtttt cacaggaaca aaaattattt 103140
atgttttcat atttcattga tttgcctaat gtttaatgaa gttagataat aataataagt 103200

acaactggca taaacagatt tgggaccaga tactactgac atttgataag cactcaactc 103260
ttcggagcga agtgtctggg caaatgagac atctttgtag tttagaatgc tcaaggagaa 103320
ctagccatca atctgtatat tttagggata tggaaaacat aggttaatat ggtaagtggc 103380
aaaatggagt ttggttaaga gaactttcgt tctgtttgtc taaggtttaa tgtttaaagt 103440
tgaacttacc tgcaatttgg gtatattatg tgtgtatctt agagggcttt gctaatacat 103500
tgatagtggg gtataattga caagggacca ttcattattc cttttcttcc cagtcctaca 103560
tttatgttaa tgcattgact tggttttccc ttaactccac agcaacttcc tattattctc 103620
aatactaagt caccactgaa aggggtactg atggtggtgg tctccacttt ccctcttatg 103680
tctggcattt tccttagttg gaagtctcag gagagggacc catgttattg taagtcctct 103740
atggaattgt gttccatggg attctgtgtt aattgtctgt ccacttgaat tcactttgct 103800
tatcagcaaa ttttgtcttc tttatatata tatttttct gattctctcc cttatttctt 103860
ttgttttttt gctgcactaa ctaaagcctt ttctatgttt ttgcaggctc tttcttgata 103920
aggatattcc taggtattgt ctgtttgctt ttttgcacgt aggattctag tttatcagaa 103980
tgctttctgt cttatgttgc tttgccattt taattttgta gtcccctccc catgtaaacc 104040
tactgtacct aggatagagt agaccctgtt ctgtacccta agaagcgtat agcttcttaa 104100
atgatttta gagtttcaaa ttatttttc attccatttg ctaaacttgt cagttttatt 104160
cttagactct ttcaagtttt cctcttctac attacaaagt agaatgtttc tcccttcatt 104220
tcatacagaa tctcaggttt tgttttctgt tttcctcttt atctgtcttc tctccgtcaa 104280
aagcatgaag ggtttattgc tttataaaac ctgagtccat tgcagtcttg caaaactgtt 104340
cactttcttc tctgttttgt cctgtgtggt aatagagaca aatttggcta yttggatagc 104400
tgagtctatg agaaatattt tatatatatt ttttaacatt tatctttctc ctgtcagggc 104460
tgcttgtttt atctgtttat tttcttcttt tggaatagga aaaatcctgg tttggtatta 104520
aaggaagttt tttaaaaaat ctgtctcatt tagggaagaa aattgaagga agtgatacat 104580
gacaatataa cctaaaatgt taaaaaaaat acaacataaa cgggctgggc tcagtggctc 104640
acacctgtaa tcccagcact ttgggaggcc aaggcgggtg gatcacgagg tcaggagttc 104700
aagaccagcc tggccaagat ggtgaaaccc catctttact aaaaatacaa aaattagcct 104760
ggcatgatgg tgggcgcctg taatcccagc tactcaggag gctgaggcag aggatcgctt 104820
gaacccggga ggcggaggtt gcagtgagcc gagatcgcgc cactgcattc cagcctgggc 104880
gacagagcga gactctgtct cgaaaaaaaa atacataaac atatttatat atttgactga 104940
tgttactgat ataatgattt tatccctcaa tttggacccc attcagttta catttcagag 105000
attctcctaa tgaactcaaa agggttcatc agatgtgacc atgatataag aaggtggctt 105060
ctgttttgtt ttcttgagac ggggtgttgc tttgttgccc aggctagagt gtagtagcta 105120
tacctgcaga tgctttaaaa aaagaaaaaa aaatgcggcc gggcacggtg gctcacaccc 105180
gtaatcccag cactttggga gaccgaggcg ggcggatcat gaggtcaaga gatcgagacc 105240
atcctggcca atatggtgaa gccccatgtc tactaaaaat agaaaaatta gctgggcgtg 105300
gtagcatgtg cctgtagtcc cagctactca ggaggctgag gcaggagaat cgcttgaacc 105360
tgggaggcgg aggttgcagt gagccaagat cgtgccacta cactccagcc tggcgacaga 105420
gctagactcc gtctctaaat aaataaataa ataaataaat aaataaagtg cggtagcatg 105480
atcatagctc actgcaacct tgaattcctg ggctcaagca gtcctccctc ctcagcctcc 105540
cgaataggtg ggactccaga cacgtgccac catgcctagc taatttaaaa ttttttttgta 105600
gagacagggt ctcgctatat tgcccacgct ggcctcaaat tcctggcctc aagtgatctt 105660
cccactttta tctcccatag tgctgggatt acaggtgtga gctaccacaa ctgaccagaa 105720
gttggttttt aattgttcat aaataataaa ttcaaaggaa tatatccttc aaagcagtca 105780
ccttgaaacc tgtttttcct ctgatgctgc tgctgttcag aacatcttga gaaatccatt 105840
tggaattact tctcaccacc ttaggaggca tacaggaaaa tatgtcttgg gattcagccc 105900
tgtttgtaac cctaaattct attgccagct tgattaaatt gacatttctc cccagtcttg 105960
gttgtatttc ttacaaccac ctctcaagaa ttgccagtgt aatggctcac acctgtaata 106020
ctagcacttt cctactaagg taggaagatc acttgagccc aagagttcaa gacctgcctg 106080
ggcaatgtag aacttgtctc taaagaaaaa aaaaaattaa ttagccgggt gtggtggctc 106140
gtgcctatat tcccagctac tcaggaggct gaggtgggaa gatcacttga gcccaggagt 106200
tagattgcac cagtgcactc cagcctgggc aacagagcaa gacccttgtc taaaaataaa 106260
ataggccagg cgcagtggct cacacctgta atcccagcac tttgggaggc cgaggcgggt 106320
ggatcacctg aggtcaggag ttcgagacca gcctggccaa catggtgaaa ccccatctct 106380
actaaaaata aaaaaattag ccagacttgg tggtgggcgc ctgtaatccc agctacttgg 106440
gaggctgagg caggagaatc gcttaaaccc gggaggcgga ggttgcagtg agccaagatc 106500
gcaccattgc actccagcct gggtgacaag agtgagactc tgtctcagaa aaataaataa 106560
ataaaatata tatttttta aattgaatta ccacctttgg tgggtattct aacagaactt 106620
gctgcagcaa gtccaaagca aaaaaaaaaa aaaaagaaag aaacttttt tgacaaaaga 106680
aaaaaatgaa agaaaaattt ttccctcaat acaggaagac tttagtggag tcaaaatttc 106740
ctttgtgtta accaggtttc tgcaagtgaa attttctctt tatgttaagt gaaagcttag 106800
cctgaaaaca gaaatgtgaa gaggcttttt tagacaagtt caaaacattt gcctcagttg 106860
aaagagtcag tgtcatgaaa cctcagcagt ggtcctggga ggtgccacgt tatggtttgg 106920
aacttaatga aggcagcagc attgatagaa atttctgtca aatatcacca aagtcataaa 106980

tgccgtcact ctaagaagag gggtttattt aactttagca tggatgtatc cagcctctga 107040
aatgggacaa ataagccccg ccccttctag gtactttgac aaaaacagga acacaataaa 107100
agagcagacc tctggaggct gaggcaggag aatcacttga acccggcatg aggaggttgc 107160
agtgagccaa gatcatgcca ccctactcca gcctgggtga caaaagcaag actctgtctc 107220
aaaaaaaaaa aaaaaaaaca gatctcacaa tgatacagtc agttctcaag tggcatctgt 107280
tgaagcagtg gatgtcgtca cccagtatta gaatatgtct taaaccaaat taaaacttta 107340
ctttacacag agttttatgg acacatgaag ggtttcctcc tgtgcccagg gagcccaaga 107400
actatagaaa ggccttcatt gtccatcatt tctgttcctg tttgggtcac ttggtggctt 107460
tgtctttgac ttgttagttg tttctctatc ttctgtgtct ttgtagagaa ttttcttgac 107520
ccacactcac tgcctcaaag actagacatt tgagagagta ttaagtactt gaagaagact 107580
tagatctaga agcagctatg ttagatgatc agacctcata cactgccact tggaagatga 107640
ggtgtggtca tggggttcag tttcagaaat gtgttctctc tgttgatcgt ttctggagga 107700
agaaacatgg ggatgtgttt agagggagga tatgccagct acttatcata ggaagtctag 107760
aaaaggcagc catagtgagg tcactatttg aaaaggagta agaaggaatg tattgggcag 107820
gaaagagtgt gcatagactc ttctaggtga atcaggtgag ctgtgttcta tctgcggaga 107880
actagtgatg gtggggcaga ccaggcatct tcccctccac ttgtcatagg ggtctgagaa 107940
gggaaacatg gggcagaatt gggaatcaat ctgtaacact gcttttaaac taagtgctgt 108000
atttcacatt agacttgcag aactggaaga tttaatgtca attagagtta atgttgtttt 108060
accagtagaa aaaaaacgat attagaacaa taacaacaaa atagtttgtg ggtatctatt 108120
aaattaagtt tccatttgtt acataaatgt aagattaatt tttcaaatgg ctttttttt 108180
taagctaaca aatggtattt gtctggctct agtgttaaca atttattctc ttttactgaa 108240
acatcagtta agaataacct tgtattcttt tatttaaaat aaggctctcg gccggcacag 108300
tggctcacac ctataatccc agcactttgg gaggccaagg tgggcagagc gcttgaggtc 108360
aggagtttga gaccagcctg gccaacatgg tgaaacccct cctctactaa aaatagaaaa 108420
attagctggc gtagtggtgc gtgcctataa tcccagctac tcaggaggct gagtcaagag 108480
aatcgcttga acccaagagg cagaggttgc agtgagctga gattgcgcca ctgcactcca 108540
gcctgggtga cagagtgaaa ctccgtctca aaaataaaat aaaataaaaa aataataaaa 108600
taaaataaaa taatgctctc tatacttaaa cccatctacc cagtttctac cattaatccc 108660
ttcagatata tctgtggtgt ttctgaggcc cagcagcact ggattgaatg atgtttagcc 108720
ctcttgtgat gtgaagattg aagctgacaa gttaccaagc attcacttcc tgctgtcatc 108780
attaactgac aggagagctt acatagtctt tctttcttct ttcagtttct gtccatactc 108840
acttatccat ttcatcttaa tccctttgat tgaccttact atttttatgt ttataggcag 108900
aggcaggctt gtcaaaacaa aattaaaaaa agcacaactg gccaggcaca gtggttcacg 108960
cctgtaatcc cagcactttg ggaggccgag gtgggcagat catctgaggt caggagttcg 109020
agaccagcct gaccaacatg gtgaaaccct gtctctacta aaaatacaaa aattagccag 109080
gcgtggtggc gggcgcctgt aatcccagct actcggaagg ctgaggcagg agaatcactt 109140
gaacccggga ggcagaggtt gcagtgatcc gagattgtgc cactgcactc cagtatgggc 109200
aacagagcaa aattctgtct caaaaaaaaa aaaaaaaaag cacaactaag tgttccttaa 109260
tgatgggtgg ggttgggagc aagaggatgg ggacctatct tttcctccta ggagataagg 109320
gttcctctct gttattagag tttgataaac tgaaagtgtc tgttggttgg cgtatggttc 109380
tttaggatga ttgacagtgg cagtacctgg actaacgtga ctttagcttc ttttggatga 109440
ttccctcttg tctttctgac cttaggacct tctaccgctt tgaggctgtg tgggatagct 109500
ctctgcataa ctcccttctt ctgaaccgag tgacacccta tggagaaaag atctacatga 109560
ccttgtcggc ctacctagag gtgaggagac ttggaacttc agttgatgcc aacagtcagc 109620
cctgaacagt gtgcaggtta cagcaggaag tggtcagcga agattccact tgttttgtcc 109680
tctttcttcc tccacatcct taagtttttc attttattat acattgctaa ctcttctgat 109740
tatattagac tctattcctg ggttctgttt actcaaaaaa tcacatgact cattgcagct 109800
tcagctttag tagcctaatc accattgttt gaagaatgtg tgatttttaa gaaaatgtca 109860
cctcttgttt ttgtagttgt ggacagagtg taatagtggt taatagccca gacccaagac 109920
ccagcctgtc ggttcaaatc cctgtgctgc cacttacttc ctctatgtcc ttggagaaac 109980
tctctttgcc tcagtttcct cacctatcag gagaaacaac atttacctca gataaaaata 110040
aatgcatctc taaggagtac tacctactct tgcacatatt atccattggg tggctgttga 110100
aataattggt ggagctaaaa tccttcagtt ctcacaactg ggactcgctt tccaaatacg 110160
ttcgtgtata gactctagtc acagtactct gcctgcactt catcagcacc taccctgtct 110220
gctttcagct ggatcattgc atccagccgg ctgtcatcac caaggatgtg tgcatggtct 110280
tctactcccg agatgccaag atctcaccac cacgctctct gcgtagcctc tttggcagcg 110340
gctactcaaa gtcaccagat tcgtaagttt ttcacacaag ttagcttcca gtgtgtttgt 110400
tcagtacaac agaatcatta gtccttaagt attctcggct tactgtctca cggttagttt 110460
acagtttatg aaccattttg gtaactgaca cctggaatgt actgacttgc atgctgacat 110520
ttaataaaat gttgatggat ggacgagtta tgagggaagg aaggaatatt acattgtgaa 110580
cacgtctaaa acaattttgc aaagactaaa tttgcaaaga cttttaaatg tcatcatcat 110640
tatttctaac agagtaccat tgaaattttc ttctttttaa gatttgttat gcatatttaa 110700
ggcatacaaa atatgtcagg aataatacag tgaaggccat acacagtggc tcatgtctgt 110760

aatcccagca cgttgggagg ctgaggacag agaattgctt aagcctagga gtttgagacc 110820
agcctgggca acacagaccc catctccaca gaaaatttta aaattaccca ggtgtggtga 110880
cacacacctg tagtcccagc tactcaggag gctgaggtgg gaggatcact tgagcttggg 110940
aggttgaggt tgcagtgagc catgatagag ccactgcact ccagcctggg tgaaacagtg 111000
agaccctgtc tcaaaattta aaaaaaagaa aagaaaaagg aaaaaaaaaa aaaagaaaga 111060
aacaatacaa tgaatacctg tgtatcaccg cccagcttaa aaataagtgt tactaataca 111120
atagaaggta ttatacctcc tactactcac aatccccaga agttgttgat tatcatttca 111180
ttctagtcat tgtttccctc atatcgtctc tcttatagta agaaaagaca aatttttagt 111240
tttcacattt atattttaaa ttttctagtt tatcattcaa ttttattcac cctattgaaa 111300
taaagtgttt aaaagataga gatgaggtgg ggcatggtgg cttatgccta taatctcagc 111360
acttttgat gccaaggtgg aaggatcact tgaggccagg agttcaagac cagcctgggc 111420
aacataagga gaccctatct ctacaaaaaa acttaaaaat tagctgggta tagtggtgta 111480
tgcctgaagt cccaactact caggaggctg aggcagaagg atcacattag tccaggagtt 111540
cgaggctgca gtaagttgtg attacatcac tgcactccag cctgggtgac acagggaaac 111600
cctatctata aaaaaattaa attgagatga aagcggactc ctctaggaaa aggtatgcta 111660
cttaggttca aaataagaaa agagcggtga caaataagaa atgcctggaa ttatatttga 111720
gtccctgcag agaagacact attttaagta gcagaaattg aaattttcct gtttttgtgc 111780
tttaagagat taaacaattg ttttattttc ttcaaatagg aatcgagtca ctggcattta 111840
cgaactcagc ttatgcaaaa tgtcagacac aggtagtcca ggtaagctct tgtggattga 111900
ggaggtgata gttatctttg tgtatgtttc aagtatcctg tatttaggct gggcacggtg 111960
gcacacacct gtaatcccag cgctttggga ggccaaggcg ggtggatcac ctgagtccag 112020
gagttcaaga gcagcctgac caacatggtg aagccccccc ttctaccaaa aatacaaaaa 112080
ttagccaggt gtggtggcgc gtgcctgtaa tcccagctac tcaggaggct gaggcatgag 112140
aattgcttga acctgggagg cagaggttac agtgagccga gattgtgcca ctgcactgca 112200
gcctaggtga cagagtgaga ctcagtctca aaaaagaaaa agaaaaaaaa atcaaattat 112260
taagtcagat ttgttctgta ttctttttaa tgtgttttta aactgctctt gttcctctag 112320
ctaggcggag ttaagcagca cttagcattt ctatgtttgt aaaccgaaac atgaacaaaa 112380
gttaaggttt tctaatagca tttgattaat gtgagtttga gtcattttac agaatataca 112440
tggaaaacaa gtcagggccc acaggcaaga cttagataag taaagaagac tttgagtatg 112500
aatccaagta gcacctaatt attatactat ccagatagtt tcttgaggaa ttttttaat 112560
taatataaat tcatgtattt aatatgtaaa accagcataa tactaaaaaa aattaacagg 112620
aaataataga ggctaaagtt ctctgtcatc tgtcagccag tgagctactg agtgtgtgta 112680
tgtgtgtgtg tattcccttc cgtcttttta aaaatttta tttttaggga caggatcttt 112740
ttttttttt ttttttttg agacagagtc tcgctgtgtc acccaggctg gagtgtggtg 112800
gcgccatctt ggctcactgc aacctctgcc tcctgggttc aagtgattct cctgcctcag 112860
cctcctgagt agctgggttt gcaagcaccc accaccatgc ccagctaatt tttgtatttt 112920
tagtagagat ggggtttcac tatgttggcc aggctggtct cgaactcctg acctcagatg 112980
atagcccacc tcagcctccc aaagtgctgg gattacaggc gtgagccatc acgcccggcc 113040
ccatagtgta tagactacta cagactattt aatatttctg cattttcacc taacatcata 113100
attttatcat tttcttaaat tgcatcccag tcttcataat tatcaaagaa tgtaaaactc 113160
agtttttaaa aacgttcagt tgctttgcta tttaaacaaa atcccatata tcttcagacc 113220
aaactgtctt agtttgctac gtataagaat tattaattcg tggccaggcg cggtggctca 113280
cgcctgtaat tccagcactt tgggaggctg aggcaggtgg atcacgaggt caggagatcg 113340
agaccatcct ggctaacaca gtgaaacccg gtctctacta aaaatacaaa aaattagctg 113400
ggcatggtgg cgggcacctg tagtctcagc tactcgggag gctgaggcag gagaatggca 113460
tgaacccagg aggcggagct tgcagtgggc tgagatcacg ccactgcact ccagcctggg 113520
caacagagcg agactccatc tcaaaaaaaa aaaaaaaaga attattaatt cgttttgacc 113580
taaacttgga attgaatttt ttttctgaaa caaaaacttg tttcctcttg tcttaggtat 113640
gcagagaagg agaagaaaaa tcttagatac gtcagtggca tatgtgcggg gagaagagaa 113700
cttagcaggc tggcggcccc gtggagacag cctcatcctt gagcaccagt gggagctgga 113760
gaagctggag ctcctacatg aggtatccag gggcagggtt gttcagatgc aagaactctc 113820
ggacaagatt gccaaagtat cagtcttcct ccccgctgct gcatgtgatc atagcttttc 113880
acttttctct cctgaggtct taacgagctt tgtgtttgct atagcagtag tattgatctt 113940
ctcaggtgga aaaaacccgc cacttttttgc tgctgcgtga gagacttggt gacagcatcc 114000
ccaaatccct gagcgactcg ttatccccca gcctcagcag tgggaccctc agcacctcca 114060
ccagtatctc ctctcagatc tcaaccacta cctttgaaag cgccatcaca cctagcgaga 114120
gcagtggcta tgattcagga gacatcgaaa gcctggtgga ccgagagaaa gagctggcta 114180
ccaaggtgtg aatcccttcc tctttgctga acgtcttccc acaaggctcc acaaactagc 114240
ctctcggttt attcattttc aacacctttg ttcgaggtgt ttgaaggcct gtgataatac 114300
tgtgaatgta gaaataaaaa gacgcagttc ctaccctcaa caagtttaca gggccaggca 114360
cagtgcctga tgcctattat cccaccactt tgggaggctg aagcaaaggg attgcttgag 114420
cccaggagtt aaagacaagc ccaggctaac atagtgagac cttgtctcta caaaaacaaa 114480
aattaggcca ggcatggtgg ctcacacctg taatcccagc actttgcagg gctgaggcag 114540

gtggattacc tgaggtctgg agttcgagac cagcctggcc aacatggtga aaccctgtct 114600
ctactaaaaa tacaaaaatt agccaggcgt ggtggcgagc gcctgtaatc ccagctactc 114660
aggaggctga ggcaggagaa ttgcttgaac ccaggaggcg gagattgcag tgagccgaga 114720
tagctccact gtgctccagc ctgagcgaca gagtgacact ccatctcaaa aaataaaata 114780
aaataataaa aaattaaaaa attaaaaaat ataaaggttg cagaatatta ggtcacaagc 114840
atataatага gatgtgtgtt tcaatgctat gggagtcaga ggagaagtgt ctatggtggg 114900
gagctcatga gctgaccttg aaaggatgcg caggagggag gcagcagtgt cgacacctgc 114960
gtgatcagga gggaggcagt ggtgtcctgc gtgatcagga gggaggcagc ggtgtcttgc 115020
gtgatcagga gggaggcagc ggtgtcctgc gtgatcagga gggaggcagt ggtgtcctgc 115080
atgatcagtc ttggggaatt cacgtgggac tggaacagga ggtgtggagt gaagaatctt 115140
gtatactgca ctgaagagtt tagagtggat gatgggagca aggcagaaat atgttaaaca 115200
ggtatttgtg ttttagaaag atttctggcg aggggcggtg gctcacgcct gtaatcccac 115260
cactttgaga ggccgaggca ggtggatcac aaggtcagga gattgagacc atcctggcca 115320
acatggtgaa accccgtgtc tactaaaaat acaaaaatta gccgaacgtg gtggcgtgcg 115380
cctgtagtcc cagctactaa ggaggctgag gcaggagaat cgcttgaacc caggtggcgg 115440
aggctgcagt gagccgagat cgcgccactg cactccagcc tgggcaacag agtgagacac 115500
catctcaaaa aatatatata tttctctggc agcaataaag agaatggatt gatggtgaag 115560
agatacactg aggaagggag aattctttat ctttattaca gtaatctgat gagaaatgat 115620
aaaagctttt taaaatttta tttttttga gatggagtct cactcctgtc acacaggctg 115680
ggatgcagtg gcatgatctt ggctcactgc aacctccact tcccaggttc aagcgattct 115740
ccttcctcag cctcccaagt agcttggagt acaggcatgt gccaccacac ctggctaatt 115800
tttgtatttt tagtagagac agagtttcac catgttggcc aggctggtct tgaactcctg 115860
acctcaggtg atccacccgc ctcagcctcc caaagtgcta ggattacagg tgtgtggacg 115920
cccagccaga taaaagtttt ttaaagaagg cagaggtggc tgggtgcggt ggcacacgcc 115980
tgtaatccca gcactttggg aggctgaggt gggcggattg tgaggtcaag agatcactat 116040
gatgttttta ttttttattt ttatctttat taaatttttt tttttttga gatggagtct 116100
tgctctgtgg cccacgctgg agtgcagtgg tgcgatctcg gctcactgca gccttcactt 116160
cccaggttca agcggttctt ctgccttagc ctcctgagta gctgggatta caggtgtgca 116220
ccaccatgcc cagctcattt tttgtaggtg ttttttttgt ttgtttgttt gtttttgga 116280
gaggggatgg agtcttactc tgctgccagg ctagagtgca gtggtatgat ctcagctcac 116340
tgcaacctcc acttcccggg ttcaagtgat tctccttcct cagcctcccg agtagctggg 116400
attacaggtg cccaccatga cgcccagcta attttcttat ttttagtagt gatggggttt 116460
caccatgttg gccaggatga tctcgatctc ctgacctcgt gatccacctg cctcagcctc 116520
ccaaagtgct gggattacag gtatgagcca tcacacccag ccagaaccat ggtgttttta 116580
gtaattcact ttggtaatct tgagaaaggt ttttttaagc cacttgcttt atctttatta 116640
gacatacata gaacttccct tactatatac ttagttctaa tttctagagc tgtttgggca 116700
acatcggagc tgcactgtgg gtttaggaca tgttatttcc cctcctgcct tgtatccagg 116760
gcttcattta gtcattgagg tttttgtttg tttgttttta aatttactga tttattagaa 116820
aggacagtgc aaaggagacc aatgaacggc agatggacga atgcataggg agaggagatg 116880
tggagcttcc atgccctccc cgggaactca gcccttcagg agcctccacg tggtcagcta 116940
tcccaaggct cctggctgtt gacattttat gttcagcttg cacttctctt tctcttcttt 117000
agtgcctgca acttctcacc cacactttca acagagaatt cagccaggtg cacggcagcg 117060
tcagtgactg taaggtgagc acattgactg taatttttag ccagtatgtt gataactgat 117120
ttctccacag cagcccagat tacctattcc tggtttttgc tgcttttaag caacttgtca 117180
tgggcatagc attgtatttg aaaatttatg acatactgct ctggtattca ttctaatttt 117240
tcagagtccg aacactgact tctgaagata aaagtactcc tttgtgtctc ttagagtgat 117300
tatcagatgg gaaacatttt ggctttttca tgactccttt ggaggagaat attctatggg 117360
gaggtggtat gttattcttt gccagggtac aaggaaaccc tgaggttcct ggtggcataa 117420
agttttattr acttcaacaa agagtgaagt aaacacttca gagaatttct gtgttattca 117480
ctcaattcta gtcagcttga ccttaaaccc tcccggctca ttcctgcctt ggggtctttg 117540
cacacctact atttctttgc ctggaattgc ctttctccag gtattgccat ggttggctcc 117600
cttacctctt tcatttttct actccattat cacctcctgt gagtcgactc tgttaacacc 117660
tgataataaa tctgacagct gtctgaaact ttctacctac ccctctttcc tgctttattt 117720
tttgctatag cactaattac tacctgatgc tttttaaagt gtttgttgag tggtttattg 117780
tctgtcttaa gctccaggga ggcagatcct ctgccatatt attagaacaa tgcctagtac 117840
atggaaggtg caccactgta ataccactta tattaattca ggggctgggt gtggtggtgc 117900
atacctgtaa tcccagcact ttgggaggcc aagatgaggg gatcgcttga ggccaatagg 117960
ttgagaccag cctgggcaat atagcacaaa taccctatct ctacaaaaaa atttttgaaa 118020
aattagccaa atgtggtggt acacacctat agtcctagct actcaggagg ccaaggcagg 118080
attgcctgag cccaggagtt tgaggctgca gtgaggtatg attgtgccat tgcaccccag 118140
cctgtgtgac agtatttttt tgtctcttaa tgataacgga ctgggggct ggctgcggtg 118200
gctcatgcct gtaatcccaa cactttggga ggccaaggcg ggaagatcac ctgaggtcgg 118260
gagtttgcga ccagcctgac caacatggag aaaccccatc tctactaaaa atacaaaatt 118320

agccggatat ggtggcgcat gcctgtaata ccagctactc aggaggctga ggcaggagaa 118380
tcgcttgaac ctgggaggtg gaggttgcag tgagccaaga tcgtgccatt gcactccagc 118440
ctgggcaaca agagtgaaac tccgcctcaa aaagaaaaca aaaaacaaaa aaacccccaa 118500
aataatgggc caggcatggc ggctcatgcc tgtaatccca acactttggg aggccgaggt 118560
gggcggatca cctgaggttg ggagtttgag accagcctga ccaacatgga gaaaccctgt 118620
ctctgctaaa aatacaaaat tagccgggcg tggtggcaca tgcctgtaat cccagctact 118680
caggaggctg tggcagaatt gcttgaacct gggaggtgga ggttgcggtg agccaagatc 118740
gcaccactgc actccagtct gggtgacaga gcgagacttc gtctctccaa aaataaataa 118800
agataatgca ggagaggcca gtgacagtgg ctcacatctg taatcctgcc agcttgggag 118860
gtcaaggtcg atggatctct tgaggccaag agttcaagac gagcctgggc aacatgacaa 118920
aaaccccatc tccacaaaaa aatacaaaaa ttagttgggt gtggtggtgt acacctgtag 118980
tcccagctac tcaggaagct gagatgggag gatcacttga gtcctggagg ctgaggctac 119040
agtgagccaa catcgcacca ctgcactaca gcctggtaga catagcaaga ccttatctcg 119100
aaaaagaaaa taataataat aataataaga agaagaagaa gaagaagaag acatagcaag 119160
accttatctc gagaaagaaa ataataataa taataataat aataatgaca acaataataa 119220
taatgcagga gggcagtagg aacataaaaa accaaccttg ccccttttat aaatgtcttg 119280
attcctggtg tggcacttag gccgtaaaat atttgtaggc ttgtttttttt tctttttga 119340
aaatattttc accagttgtg gtggtacacg cttgtagtcc caggtactca ggaagccagg 119400
gcacaggagg atcgctggag cccaggagtt ggaggccaca ggaagctagg attgcatcac 119460
tgccctccag cctgggcaat agagcaaggc cttgtctctt aaaattttac ctatgtattt 119520
atttattgat aaattatgta taggcttgtt ttcttaaatg accttctgaa acaagatgta 119580
tcaaagtgag gttctggttg ttgaaactcc ctattgttac tgtagaggca gctttttca 119640
gctgaatgta acttgtagtg ttcggtttgc ttccagttgt ctgatatctc tccaattgga 119700
cgggatccct ctgagtccag tttcagcagt gccaccctca ctccctcctc cacctgtccc 119760
tctctggtag actctaggag caactctctg gatcagaagt aagtacccag atttcactga 119820
gagaagtcaa tctaagaacc aaggtaaatg tcaaccttcc tctagctcaa tggttcttga 119880
ctgaaggcag caccttctcc ctaggcagca tttggaaatg tgtgggagcc tttttggcta 119940
tcacagtgac tggagaggtg ctagctactg gcatctcttc acatcatagc ttactaacct 120000
tctgcatcat caactcatca tgactgtctt cttcatccct ttctgctagt tttatcctcc 120060
attcctaccc ctgcttgtct ccccagtcat aataccagac ataggaagtg ctatcttggg 120120
aaaaatattt tgaaactgtt ttggaaaaaa atattttgat tccacagcaa attatgtagg 120180
taagaaaaat aaagacaagg ccaggtgagg tggctcccag cactttcgga tgccaaggca 120240
ggaggattgc ttgaggctgg gagttcgaga ccaggctggg caacatagcg atatgctgtc 120300
tctacaaaaa ataaataaga tacccaggcg tggtggtctg tgcctgtggt cccagctact 120360
cgggaggctg aagtgggagg atcgcttgaa tccaggagtt caaggttgca gtgaggtttt 120420
attattgcac cactgcattc caccctgggc gacagagcaa gaccctgtcc ctaaaaaaaa 120480
taaaaaaaag acaaatgtag tctagagtcc atgattcatc atcataatcc gttctttgca 120540
acacaccctc ttgtgcatct ccatcaatta tgctcacctg gttatcatcc acctgtctat 120600
cttattgtac caacactgac acaggcaaag gtggctggag aaaataacac aaccaggtaa 120660
tttcattaaa tttaggatgc cacagactgt agggactatt ctgtgccact aagagaagaa 120720
gaaaaagcac tgacaactaa cttaagcatg ccatcaatta taagatgaat cccaatttaa 120780
gaaatggtgg ctgagtgcag tggctcatac ctgtaatccc agcactttca ggggctgagg 120840
tgggaggatt gcttcagccc acaagtttga gaccagcctg ggcaacatag ggagaccccc 120900
atctctacaa aaaataaata aaatagttgg acatgatggc atgcgcctgc ggttccagct 120960
actcaggagg ctcacatggg aggattgctt gagcccggga ggttgaggct gcaatgagcc 121020
atggttgcac cattgcactc cagcctgggt gacagagcaa gaaatcagcc aggcataggg 121080
gcatgtacct gtaatcccag ctactgggga ggctgaggca ggaagattgg ttgaacctgg 121140
gaggcggagg ttgcagtgag cc 121162

<210> 2
<211> 10884
<212> DNA
<213> Homo sapiens

<220>
<221> exon
<222> 1601..1750
<223> exon

<220>
<221> exon
<222> 2139..2331
<223> exon

<220>
<221> exon
<222> 2540..2658
<223> exon

<220>
<221> exon
<222> 3830..8884
<223> exon

<220>
<221> misc_feature
<222> 8885..10884
<223> 3'regulatory region

<220>
<221> allele
<222> 2397
<223> 12-602-196 : polymorphic base C or T

<220>
<221> allele
<222> 2551
<223> 12-602-350 : polymorphic base A or C

<220>
<221> allele
<222> 4500
<223> 12-587-379 : polymorphic base A or C

<220>
<221> allele
<222> 6119
<223> 12-596-124 : polymorphic base A or G

<220>
<221> allele
<222> 10130
<223> 12-808-52 : polymorphic base A or G

<220>
<221> allele
<222> 10153
<223> 12-808-75 : polymorphic base G or C

<220>
<221> primer_bind
<222> 2203..2221
<223> 12-602.pu

<220>
<221> primer_bind
<222> 2600..2620
<223> 12-602.rp complement

<220>
<221> primer_bind
<222> 4479..4499
<223> 12-587.rp

<220>
<221> primer_bind

<222> 4858..4878
<223> 12-587.pu complement

<220>
<221> primer_bind
<222> 5996..6015
<223> 12-596.pu

<220>
<221> primer_bind
<222> 6423..6443
<223> 12-596.rp complement

<220>
<221> primer_bind
<222> 10079..10098
<223> 12-808.pu

<220>
<221> primer_bind
<222> 10523..10543
<223> 12-808.rp complement

<220>
<221> primer_bind
<222> 2378..2396
<223> 12-602-196.mis

<220>
<221> primer_bind
<222> 2398..2416
<223> 12-602-196.mis complement

<220>
<221> primer_bind
<222> 2532..2550
<223> 12-602-350.mis

<220>
<221> primer_bind
<222> 2552..2570
<223> 12-602-350.mis complement

<220>
<221> primer_bind
<222> 4481..4499
<223> 12-587-379.mis

<220>
<221> primer_bind
<222> 4501..4519
<223> 12-587-379.mis complement

<220>
<221> primer_bind
<222> 6100..6118
<223> 12-596-124.mis

<220>
<221> primer_bind
<222> 6120..6138
<223> 12-596-124.mis complement

<220>
<221> primer_bind
<222> 10111..10129
<223> 12-808-52.mis

<220>
<221> primer_bind
<222> 10131..10149
<223> 12-808-52.mis complement

<220>
<221> primer_bind
<222> 10134..10152
<223> 12-808-75.mis

<220>
<221> primer_bind
<222> 10154..10172
<223> 12-808-75.mis complement

<220>
<221> misc_binding
<222> 2385..2409
<223> 12-602-196.probe

<220>
<221> misc_binding
<222> 2539..2563
<223> 12-602-350.probe

<220>
<221> misc_binding
<222> 4488..4512
<223> 12-587-379.probe

<220>
<221> misc_binding
<222> 6107..6131
<223> 12-596-124.probe

<220>
<221> misc_binding
<222> 10118..10142
<223> 12-808-52.probe

<220>
<221> misc_binding
<222> 10141..10165
<223> 12-808-75.probe

<400> 2
gcctcccagg ttcacgccat tctcctgctt cagcctcccg agtagctggg actacaggtg 60
cccgccacca cacccagcta attttttatt ttattttatt tatttattta cttttgagat 120
ggagttttgc tctcgttgcc cacgccagag tgtagtggtg caatcttggc tcactgcaac 180
ctccaccttc cagcttcaag tgattctcct gcctcagcct ccaaagtagc tgggattaca 240
ggcacccgcc accatgccca gctatttttt tttttatttt tagtagagat cgggtttcac 300
cattttggtc aggctagtct caaactgctg accttatgat ccacccgcct cgacctccca 360
aagtgctggg attacaggcg tgagccactg cacccggcca cacccagcta atttttttgta 420
tttttagtag agatggggtt tcaccgtgtt agccaggatg gtcttggtct cctgacctca 480
tgatctgcca gcctcagcct cccaaagtgc tgggattaca ggcgtgagcc actgcgccgg 540
cctttttttt ttttttttga gacagagtct cactctgtcg cccaggctgg agtgctgcaa 600
cctctgcctc ccgggttcaa gtaattctcg tgcctcagcc tcctgggtag ctgggattac 660
aggggcctgc caccacgccc agctaatttt tgtattttta gtagagacgg ggtttcgcca 720

tgttggtcag gctggtctca aactcctgac ctcaggtgat ctgcctacct cggcctccca 780
aagtgctggg attacaggtg tgagccactg cacctggcta acttgatttc aaaaaaaaaa 840
aagaatgaag tatgagaaaa agttaaaatg ggctttttag atttagtgag atacaatgct 900
atgttactct acaaagatat ctagcaatac agagtatagg aatttatttc tcacacagta 960
gtcaagaggg ctggtccagg tccaggtcag tggtgcaggt cttctctgca cagccgttca 1020
ggaaccctgg gtcttcccat ttgtttgctt tttctgccat tttctaaggc agtgggcctc 1080
tgagacttca gtgtacatca tcatcccctg ggaggcttgt tgaagccccc atggctcaac 1140
cccacccaaa gtttctcgtt ccgtaggtct ggcttggagt ctaaaaattt gagtttttga 1200
acagtttccc aggtgaagct gatgctgctg gctgggagca cactttgaga aacatcgctt 1260
ttggtgtggc catgggacta aaaagaaaca tataaaggag tagagttcct attccagata 1320
atgtgttcag agggctttaa tataggcaga gagattgtgt gacccagtgg ccatcaatga 1380
tcaatttcct ttcataactt cagaaggatg acaccagttg aacgcagagt taactttctg 1440
aagccagctt gtctcctcag ctgagctctc tgcaactcaa gtttgcagct ggggtttagg 1500
gagggccatt gtgttcctcc cagtgaaaca gtactcagta tgccttgatt gtaactgatt 1560
ctcttgttac ccttttcttt ctaatctctc tattttaaag gaccccagaa gccaattccc 1620
gggcctctag tccctgccca gaatttgaac agtttcagat tgtcccagct gtggaaacac 1680
catatttggc ccgagcagga aaaaacgaat ttctcaatct tgttccagat attgaagaaa 1740
ttagaccaag gtgagtacta tattgagcag gaatgccagc tataaaaaac aaatccacag 1800
gaagaagtga ctggccagct cttccctgtg aggtctgtac tgtagtctga tgcaatctgt 1860
actgtagtct gatgcaagtt gagtctgggg tgagagggcc agcctgggtt tctccagttg 1920
ggtctcataa tgcatctgca ccctggcgca gcatgttgcc atggcacggt ttcgcttttg 1980
ccgtattact ttggcagata agattctagg ccaaataggt catttgcctg tttcatcgaa 2040
tctaaggact tggcctaagt gtggcgtatt ttgatggtcc tcagcacgat tattttcttt 2100
ttgtgagaaa ctaacttttg tcatattgtc gtttttagct cagtggtctc taagaaagga 2160
taccttcatt tcaaggagcc tctttacagt aactgggcta aacattttgt tgtcgtccgt 2220
cggccttatg tcttcatcta taacagtgac aaagaccctg tggagcgtgg aatcattaac 2280
ctgtccacag cacaggtgga gtacagtgag gaccagcagg ccatggtgaa ggtccgtcct 2340
gccctgcctt ggtttcttat tgccacgtgt gcccttctct tttgatttct gcactcyctg 2400
ttacgtggtg tgaaatttcc ctattctctc tggttttgaa cttcctttgt gaagtgctat 2460
atgccttggg aatatggcaa ggcagtcccc attgctgtct ctgtagtaac tttcttgtct 2520
acctgcattt ttctttcaga caccaaacac mtttgctgtc tgcacaaagc accgtggggt 2580
ccttttgcag gccctcaatg acaaagacat gaacgactgg ttgtatgcct tcaacccact 2640
tctagctggc acaatacggt aagaagtttt gttgttgttg ttgttgtttt tgagacggag 2700
tctcactttt tctccctgaa gtgcagtggc ttggtcttgg ctcactgcaa cctccgcccc 2760
ctggttcaag cgattctcct gactcagcct ccccagtagc tgagactaca ggcacgtgcc 2820
accattgcct ggctaatttt cgtatttttt gtagagacgg ggttttacca caatggccag 2880
gctggtctca aactcctgac ctcaagtaat ccacccacct cagcctccca aagtgctggg 2940
attacaggcc tgagccactg cacccagcct agaagttgtt ctgtgttttc tttctctctt 3000
tggcttttgg cttcaggaat atcaggataa ataggaaagg aagaattttt cttttctttt 3060
ttcttttctt gttttttttt ttttttttt tttttggaat ggagtttttg ctcttgtgtg 3120
gcccaggctg gagtgcagtg gcatgatatc ggctcaccgc aacctccacc tcccaggttc 3180
aagccattct cctgcctcag tctcctgagt agctaggatt acaggtgccc gccaccacac 3240
ccagctaatt ttttttttg tatttttagt agagatgggg tttcaccatg ttggccaggc 3300
tggtcttgaa ctcctgacct caggtgatcc accctcctcg gcctcccaaa gtgctgggat 3360
tacaagcgtg agccaccaca tctggcttga aaggaagtat ttttcaaaat aaatttaaaa 3420
gagaggaagg acaggagtgt aaattggttg tggggtggta atgaggcctc tgcatcttgt 3480
aaggcattgt catgttctga accaccaggg atactctggt cacacatgtg caccaattat 3540
tttccttctt tgtctttttg ttctcaagag ttctaaagat ggggctcagg ggagagagta 3600
aagggaggac agtgagaacc ttaatggtct cccaacctcc agccctgtga ttcagcctca 3660
cgttctgagg gcttgtgtca cagggtcccg gagtagacgg cttagaggac taggcctgcg 3720
gtgccaaata ggaagatgct ccaggaggga gcctcagagg gagaggactg tgaggggaca 3780
agacttgtac ccagacttct aatcctacct ccccgtttgt cccccatagg tcaaagcttt 3840
cccgcagatg cccgagccag tcgaaatact aagtgactct gccgagtgcc ctcactcgcc 3900
ttcgagagat aaagaaagcg ttacctctca tttctctttg tgattcttga cggtgactct 3960
tgtatgtaat cctgtggctt aactacttct ccctccttgt ccagcacttt tctagctctc 4020
ccgttcccca tctccattgc tctgtactct tttctttttt cttgtgctga gaatctcgtt 4080
agtagcatgt ggcctaacaa aaggaaaaaa tgtttttaaa cacacacaca cacacacaca 4140
cacacacaca catacacaga caaaaacaca aaaactctga ggggatctgg tgaatctcca 4200
aattattgtg ggtgtacttt ggcttccttt tgtatgatag gtccccatca tgaccacctc 4260
tgatgtctgt gctgctgtca ccaggcacct ttgtttttca agacaacata ctttttttt 4320
cttttctctg tttgtgatat cactttaatt tttcttgggt ggcttagaga ctaagggagg 4380
agacatctgg ccttttaga acctgagagg aaaaaaagag tctttttttc ccctctgtct 4440
ctttttgcca tggctaatcc ctgcatttcc attcagggaa aaggtggtag tgagcatagm 4500

actgcaacag ttatattctg agtcaaagtt ggggcttttt acggcataat tatggaattt 4560
ttatttactg gtagagagga gacgagaggc tttttcagtg ggcctgggac agtggctgct 4620
cttgactttg tgtgaaggga aatgccaagg atgcttctgg tggacttcag gggaccccag 4680
ggtttggccg tgggccgtga tggcagcagg cggtgggatg cttgtagctc ctcacagcag 4740
gattcctgcc cactgttttt tctctgttgg gagggaagct cttttctagg agtgtctcag 4800
ttctgctttt ggcattagtg atggtggtgg tacagttgga attagtgcca tgtcatacac 4860
aaatgttcca caaggcggga gtgtttcact ttctggtgat aaacttgatg gtcattgtta 4920
tgattaagat aatgccgggc aggccgggca cagtggctca cgcctgtaat ccaagcactt 4980
ggggaggccg aggcgggcag atcacgagat caggagttca agaccagcct ggccaatatg 5040
atgaaacccc gtctctacta aaaatacaaa attagtcggg tatggtggca catgcctgta 5100
attccagctg cttgggagcc tgaggcagga gaactgcttg aacccaggag gcagaggttg 5160
cagtgagcca agatcgcgct attgcactcc agcctgggtg acagagcaag actctgcctc 5220
agaaaaaaaa aataataata atgctgggta gtgaccttgt gattgttaca gctccctttg 5280
atcaaagaaa tatagctttc aggcataaac ctggaagtct ccctctgaat ccagcagttg 5340
ttttcattga tgcttgtcag gttgaagatg ctttcagtga tgctctctat actcataaat 5400
aagcaaatgt ggcaggcttt gctttctgga tcccaggatt aaaactaacc gtgaccacta 5460
ctccaaacaa aacacaatat gcctaggggc acggatgaac gtccagggag cccgggcccc 5520
aggctttgtt gcgtgttccc tgctcctctc catctggtgt ggaaacactg cccagggaga 5580
aaggaggaag ctcactgtgg acagtcttct ttccttctga cagaccaggt catctggctt 5640
ccgagatcat cagagaagat aagtctgtct ctttcagctg ccagtaagtt ttccaggatg 5700
agaggggaaa aagaaagcct ccagtgactt cagttgcttt gccagttgtc ttgggattgt 5760
tttacaccat cctttacttc ccttgctcag acctctctgt ttcaccattg ctcaggcatt 5820
caggaaagta tctgctcact cccacttggt gagtcctcgg ccttgaggtt gctgactctc 5880
aggcgttagg cagctggatg acttcccgct tcacgcagca aaggccaggg gcttgcgcgc 5940
ctctgcagag ttgttgctag ggagacttgt gtcatcatcc acaaccttgt ttctcacttc 6000
ctggttgggc tcatctctga agaacaggtc tcccagcttc gctccttatc actgcattgt 6060
gaagaggagg aaaagtgaat cacggagaga gaaaggaaag gatagaatca caggctgcrt 6120
ctgcacctga aaagtgaccc gcggaaactc tatggcggat tttttttta actttcttct 6180
tcctgttaaa acataggtca ctaactgtga tgttatttgt tttctaagtg gtatgtgaga 6240
ttttctaatg tagttagaag tttcattgtc tgatggacac aatatgccct tccggttcta 6300
ttcaaaccag caggatctgt cggtgcttag agatggctgc ctggactgga atcaaatcta 6360
atttcaggga aatgaagatg gaatttgaag gtcactttta aaattaagtc attgatgctg 6420
ctgttacaga gtgtgacaga ggatccatgt ctgtgacaca ggacggtggg aagcctgaga 6480
gagagtgaaa ttatgtgata cactgaaatg acttttgttt ttcttctaac tcatacaaaa 6540
ctggtttgga aagtctttgc tttggaagcg tcagacatta gaacaggcca aactggactg 6600
tctgttcata gcgtgcctga ataagaaggc ctcttaggga gccagaggga gcagagtggt 6660
cgtgtcctgc gtgctcttca ccctctgggg cgcccctgct gcggctggca ggtgcagaca 6720
gcctttgctg gtccccagca cgtccagggt gggtgctccc ttgcccgaca gaaccatccc 6780
cactgtgagg ctgtgagaga tttgtggcag gaactgttta tgaggctcta gttgttgctg 6840
ttgtggcggg aaagttaaga aacatagccc ttaaggaaac cacctttatg tattttctta 6900
aagcacgcct ttaaataagc aaaaacttta aaaggcagga aagagaattc ttaggcaaat 6960
tcagagaaat aagtgctagt taatactaat cacctcctcc tctgtctctc atcctccttt 7020
ctcccatcaa agcaaaatat ggcctcacca ccagccccaa atcagtgctc agaccctctc 7080
tgtgtctgtg tgccctcctg ggagtcagtc agcgctcagg ccaggactgt gcagggccag 7140
ccagcccatg cgctagtcag gagcacaggc aaggggtgct tgtggcagtg gccgggcacc 7200
tgagccccag ctcgttgtta aacgtgctga cggcaagggg caatggagtg agtttcccaa 7260
ctaagaaacc actattatat atttttttccc ttcagtcaca tagacttcag acaactctcc 7320
tatttttat ggatttttca gctcatttca gatgaaggaa ctaagtcatt gtgaactgtc 7380
tcttgagatc taaaaacaag atgacttttc ctggcacata ttccaaagca aagactttgt 7440
tgcctgctgc ttattgtcta atttacaggg atatttaatt ttgtcaggtc tatgtatatt 7500
tatccagcta tacttacttg cacagtggat tggagagaaa ggattctcca gtgtgcacac 7560
tcatcggtac tctttctgca tttccctcgt gctgtgtccc gctcgggttc caatggacag 7620
tatcagggct tgtttgactt aggtctttca gttttccttt cggttccctt ttaaaaatgt 7680
gattgttaac ctgcctcttg aaagattcaa ccgggtgtgg tggctcacgc ctgtaatccc 7740
agcacattgg gaggccaagg caggtggatc acctgaggtc agaagtttga gaccagcctg 7800
gccaacatgg tgaaaccccg tctctactaa aaatacaaaa agtagccagg cgtggtggcg 7860
tgtgcctgta gtcccagcta cttgggaggc tgaggcagga gaatagcttg aacctgggag 7920
gtggaggctg cagtgagttg agactgcacc attgcactcc agcctgggtg acaaagcagg 7980
actccgtctc aaaaaaaaag aaaagattca tgatgctgct gctcccagaa ggtttgctgg 8040
atgtgtttac ataggactct aacttgtgtg cactacagtt gttcaccagg gccagtgatt 8100
caccccagtg tgtggccaga ccatgactgt gtagcaggaa tgttttaatt tgtgcttcct 8160
tagtaaattg aaatatcagc tgagagatta tttgctgctg ttattcaaaa ggccatttat 8220
gaagttagta tttgagcccc ataagatctt taaaaagcct ccaatcattt aaaggaagaa 8280

atcagagttg ctataaaatt cagtaaaaag ctcatagcca aacggctgtg ctcagatgga 8340
aagtctgagc tgaggttggt ctcttgccaa accgtggctg ttgtgtgttg ttcttcatgt 8400
cttcgagttc attttttttc attctgccta ttctggcatc agctcacttg aggagtccct 8460
cagccttctt gtatttaagg catcgtctta gactttgtgg ctctaaagta cctgtctgtt 8520
gagatttcaa gtctcttgtc accatcctca cacatgacaa caaaacccat aatgcataag 8580
tggccttttt gaaccaagac tttgcaaact gatctctccc ccgtgaagga gttgagcaca 8640
ttagcaacaa tgtacattaa ttttggattt tcattttcat gttttatttt gtaaatatta 8700
tctgatgttt ggagcttgag tatacagact gtaaatatag ttcttgtatt tgtactaatt 8760
ctgattcttt tgctgtatag ccttagatgt gcaatgcaga cactatctaa ctgtgtgtgg 8820
taaccttgcg tcacggagct gttagtgaac gaggtaaaaa taataaaggt acagccagtg 8880
catcagaagg ttctcgatgt gcatttattc ttgcacccct tgaaaggtaa ttgcacaaag 8940
atttcttttc ttgatttgca aagataatac ctttctaaga cagaagtcac gatatcatcg 9000
cttaagtatt cctttctgat attcaaaatc gtggttttta tataaccaag aaagctaata 9060
tggtgctgtt tctttgaaag tacttttctc ccaaagatgt aagtggttat tttagctaga 9120
agatttgttg tttttctcct gaagcctcct tgtatgcctt ccaataattc tgtagtgttg 9180
tgttggcttc ttttttttttt ttttttttctt tttgagacag agtcttgctc tgtcacccat 9240
gctggagtgc agtgatgtga tctcgggtca ctgcaacctc tgcctcctgg gttcaagcga 9300
ttctcctgcc tcagactcct gagtagctgg gattacaggc gtgcgccacc atgcgtggct 9360
aatttttgta ttattattat tattattatt attattattt tatttattta tttatttatt 9420
ttgagatgga gtctcactct gtcgtccagg ctggagtaca gtggcatgat ctcggctcac 9480
tgcaactttc cacctcccgg gttcaagcga ttctccctgc ctcagcctcc caagtagctg 9540
ggattacagg catccatcac gcctggctga ttttatttttt agtagagagg ggtttcactg 9600
tgttggccag gctggtctca agctcctgac cttgtgatcc acccgcctca gcctcccaaa 9660
gtgctaggat tacaggcttg agccaccgtg cccaacctat aattatttttt ttgagatgga 9720
gtttgactct tgtcacccat gctggtgtgc aatggcacca tctcagctca ctgcaacctc 9780
tgcctcccca gttcaaacaa ttctctttcc tcagcctccc aggtagttgc aattacaggt 9840
gcccgccacc acacctggct aatttttaaa aatattttta gtagagacgg ggtttcagca 9900
cattggccag actggtctca aactcctgac ctcaggtgat ccgcccacct cggcctccca 9960
aagtgctggg attacaggcg tgagccaccg tgcctggcca tgttgggttt ttttgggtgg 10020
ggaagggtaa atggggattt gaaaagttga acatgtcaat ttaatttaac aagccctcat 10080
cacgagcaga gcatgagcag gggccactgt cctgggtgtt tggggagaar cagaagaagg 10140
gagggaggag caccctcag gttaaaagca atgatttgtt ccaaccaata tttgagtacc 10200
caccctgtgc ctggtgtgct gctgggtgca gtaattcagc agccaacaaa gcaagcccat 10260
gcacttgtgc cctcccattc cagtggggaa gacaatgttt aaacaaattt acaacatcgg 10320
tcctaagtat tttggaggaa agcagctaaa taaaaggatg aaagagggat gggaacgagg 10380
gatgtttaga tggcatagtc agggaaggcc tctcttcgga gatcacattt gagcagactc 10440
tagaatgaaa ggatagggta agccatgcaa acccctgggc aaagaggtct ccaagtagag 10500
aaatgaagcc cttgaggcag cagtgtcctc agtaggttcc agaaatagca aggtcgctgg 10560
tacagagtca gggactggga gcgcaataga agacatcaga aagtagccag ggaccaggat 10620
atttagggct taggtgagga cttggaattt cattagagtg aatgggaagc cattgaaggc 10680
tttggagcag aggagtcaca ggggcttatt tacattttta aaagaccatt ctggaaaaca 10740
gtctttgtac tcaggcaggg acaagtggaa aaaaggggca aaataagatt gttaccccag 10800
tattccaggg gaggagacgg tggctcagac taaggtagta gagtataggt gggacgaaat 10860
gggcagattc aaggtacatt ttga 10884

<210> 3
<211> 10682
<212> DNA
<213> Homo sapiens

<220>
<221> 5'UTR
<222> 1..186

<220>
<221> CDS
<222> 187..5637

<220>
<221> 3'UTR
<222> 5638..10682

<220>

<221> polyA_signal
<222> 10631..10636

<220>
<221> allele
<222> 471
<223> 12-809-119 : polymorphic base G or C

<220>
<221> allele
<222> 5487
<223> 12-602-350 : polymorphic base A or C

<220>
<221> allele
<222> 6265
<223> 12-587-379 : polymorphic base A or C

<220>
<221> allele
<222> 7887
<223> 12-596-124 : polymorphic base A or G

<220>
<221> misc_feature
<222> 2,4,5924,7850
<223> n=a, g, c or t

<400> 3

```
gngnccgctg ctcgcgctga ggtgcgtccg gtgcccggcc ccccgcgccc ccgcgcgccg      60
cggctcctgt tgacccggtc cgcccgtcgg tctgcagcgc ggctgaggaa acttggctgt     120
aacttcaaaa gaagatttga ttctttattt ctggactgca tatatatata taacaaggcc     180
attaaa atg tcg gga gcc tca gtg aag gtg gct gtc cgg gta agg ccc         228
       Met Ser Gly Ala Ser Val Lys Val Ala Val Arg Val Arg Pro
       1               5                  10
ttc aat tct cga gag acc agc aag gaa tcc aaa tgc atc att cag atg       276
Phe Asn Ser Arg Glu Thr Ser Lys Glu Ser Lys Cys Ile Ile Gln Met
15                  20                  25                  30
caa ggc aac tcg acc agt att att aac cca aag aat cca aag gaa gct       324
Gln Gly Asn Ser Thr Ser Ile Ile Asn Pro Lys Asn Pro Lys Glu Ala
                35                  40                  45
cca aag tcc ttc agc ttc gac tat tcc tac tgg tct cat acc tca ccc       372
Pro Lys Ser Phe Ser Phe Asp Tyr Ser Tyr Trp Ser His Thr Ser Pro
            50                  55                  60
gaa gat ccc tgt ttt gca tct caa aac cgt gtg tac aat gac att ggc       420
Glu Asp Pro Cys Phe Ala Ser Gln Asn Arg Val Tyr Asn Asp Ile Gly
        65                  70                  75
aag gaa atg ctc tta cac gcc ttt gag gga tat aat gtc tgt att ttt       468
Lys Glu Met Leu Leu His Ala Phe Glu Gly Tyr Asn Val Cys Ile Phe
    80                  85                  90
gcs tat ggg cag act ggt gct gga aaa tct tat aca atg atg ggt aaa       516
Ala Tyr Gly Gln Thr Gly Ala Gly Lys Ser Tyr Thr Met Met Gly Lys
95                  100                 105                 110
caa gaa gaa agc cag gct ggc atc att cca cag tta tgt gaa gaa ctt       564
Gln Glu Glu Ser Gln Ala Gly Ile Ile Pro Gln Leu Cys Glu Glu Leu
                115                 120                 125
ttt gag aaa atc aat gac aac tgt aat gaa gaa atg tct tac tct gta       612
Phe Glu Lys Ile Asn Asp Asn Cys Asn Glu Glu Met Ser Tyr Ser Val
            130                 135                 140
gag gtg agc tac atg gaa att tac tgt gaa aga gta cga gat ttg ctg       660
Glu Val Ser Tyr Met Glu Ile Tyr Cys Glu Arg Val Arg Asp Leu Leu
        145                 150                 155
aat cca aaa aac aag ggt aat ttg cgt gtg cgt gaa cac cca ctt ctt       708
```

```
Asn Pro Lys Asn Lys Gly Asn Leu Arg Val Arg Glu His Pro Leu Leu
    160                 165                 170
gga ccc tat gtg gag gat ctg tcc aag ttg gca gtt act tcc tac aca      756
Gly Pro Tyr Val Glu Asp Leu Ser Lys Leu Ala Val Thr Ser Tyr Thr
175                 180                 185                 190
gac att gct gac ctc atg gat gct ggg aac aaa gcc agg aca gtg gca      804
Asp Ile Ala Asp Leu Met Asp Ala Gly Asn Lys Ala Arg Thr Val Ala
                195                 200                 205
gct aca aac atg aat gaa aca agt agc cgt tcc cac gct gtg ttt acg      852
Ala Thr Asn Met Asn Glu Thr Ser Ser Arg Ser His Ala Val Phe Thr
            210                 215                 220
att gtt ttc acc cag aag aaa cac gat aat gag acc aac ctt tcc act      900
Ile Val Phe Thr Gln Lys Lys His Asp Asn Glu Thr Asn Leu Ser Thr
        225                 230                 235
gag aag gtc agt aaa atc agc ttg gtg gat cta gca gga agt gaa cga      948
Glu Lys Val Ser Lys Ile Ser Leu Val Asp Leu Ala Gly Ser Glu Arg
    240                 245                 250
gct gat tca act ggt gcc aaa ggg act cga tta aag gaa gga gca aat      996
Ala Asp Ser Thr Gly Ala Lys Gly Thr Arg Leu Lys Glu Gly Ala Asn
255                 260                 265                 270
att aat aag tct ctt aca act ttg ggc aaa gtc att tca gcc ttg gcc     1044
Ile Asn Lys Ser Leu Thr Thr Leu Gly Lys Val Ile Ser Ala Leu Ala
                275                 280                 285
gag gtg gat aac tgc act agc aag agt aaa aag aag aag aaa aca gat     1092
Glu Val Asp Asn Cys Thr Ser Lys Ser Lys Lys Lys Lys Lys Thr Asp
            290                 295                 300
ttt att ccc tac agg gat tct gta ctt act tgg ctc ctt cga gaa aat     1140
Phe Ile Pro Tyr Arg Asp Ser Val Leu Thr Trp Leu Leu Arg Glu Asn
        305                 310                 315
tta ggt ggc aat tct cgg act gca atg gtt gct gct ctg agc ccc gcg     1188
Leu Gly Gly Asn Ser Arg Thr Ala Met Val Ala Ala Leu Ser Pro Ala
    320                 325                 330
gat atc aac tac gat gag act ttg agc act ctg aga tat gca gat cgt     1236
Asp Ile Asn Tyr Asp Glu Thr Leu Ser Thr Leu Arg Tyr Ala Asp Arg
335                 340                 345                 350
gca aaa caa att aaa tgc aat gct gtt atc aat gag gac ccc aat gcc     1284
Ala Lys Gln Ile Lys Cys Asn Ala Val Ile Asn Glu Asp Pro Asn Ala
                355                 360                 365
aaa ctg gtt cgt gaa tta aag gag gag gtg aca cgg ctg aag gac ctt     1332
Lys Leu Val Arg Glu Leu Lys Glu Glu Val Thr Arg Leu Lys Asp Leu
            370                 375                 380
ctt cgt gct cag ggc ctg gga gat att att gat att gat cca ttg atc     1380
Leu Arg Ala Gln Gly Leu Gly Asp Ile Ile Asp Ile Asp Pro Leu Ile
        385                 390                 395
gat gat tac tct gga agt gga agc aaa tat ctg aaa gat ttt cag aac     1428
Asp Asp Tyr Ser Gly Ser Gly Ser Lys Tyr Leu Lys Asp Phe Gln Asn
    400                 405                 410
aat aag cat aga tac ttg cta gcc tct gag aat caa cgc cct ggc cat     1476
Asn Lys His Arg Tyr Leu Leu Ala Ser Glu Asn Gln Arg Pro Gly His
415                 420                 425                 430
ttt tcc aca gca tcc atg ggg tcc ctc act tca tcc cca tct tcc tgc     1524
Phe Ser Thr Ala Ser Met Gly Ser Leu Thr Ser Ser Pro Ser Ser Cys
                435                 440                 445
tca ctc agt agt cag gtg ggc ttg acg tct gtg acc agt att caa gag     1572
Ser Leu Ser Ser Gln Val Gly Leu Thr Ser Val Thr Ser Ile Gln Glu
            450                 455                 460
agg atc atg tct aca cct gga gga gag gaa gct att gaa cgt tta aag     1620
Arg Ile Met Ser Thr Pro Gly Gly Glu Glu Ala Ile Glu Arg Leu Lys
        465                 470                 475
gaa tca gag aag atc att gct gag ttg aat gaa act tgg gaa gag aag     1668
Glu Ser Glu Lys Ile Ile Ala Glu Leu Asn Glu Thr Trp Glu Glu Lys
    480                 485                 490
ctt cgt aaa aca gag gcc atc aga atg gag aga gag gct ttg ttg gct     1716
```

```
Leu Arg Lys Thr Glu Ala Ile Arg Met Glu Arg Glu Ala Leu Leu Ala
495                 500                 505                 510
gag atg gga gtt gcc att cgg gaa gat gga gga acc cta ggg gtt ttc      1764
Glu Met Gly Val Ala Ile Arg Glu Asp Gly Gly Thr Leu Gly Val Phe
                515                 520                 525
tca cct aaa aag acc cca cat ctt gtt aac ctc aat gaa gac cca cta      1812
Ser Pro Lys Lys Thr Pro His Leu Val Asn Leu Asn Glu Asp Pro Leu
            530                 535                 540
atg tct gag tgc cta ctt tat tac atc aaa gat gga att aca agg gtt      1860
Met Ser Glu Cys Leu Leu Tyr Tyr Ile Lys Asp Gly Ile Thr Arg Val
        545                 550                 555
ggc caa gca gat gct gag cgg cgc cag gac ata gtg ctg agc ggg gct      1908
Gly Gln Ala Asp Ala Glu Arg Arg Gln Asp Ile Val Leu Ser Gly Ala
    560                 565                 570
cac att aaa gaa gag cat tgt atc ttc cgg agt gag aga agc aac agc      1956
His Ile Lys Glu Glu His Cys Ile Phe Arg Ser Glu Arg Ser Asn Ser
575                 580                 585                 590
ggg gaa gtt atc gtg acc tta gag ccc tgt gag cgc tca gaa acc tac      2004
Gly Glu Val Ile Val Thr Leu Glu Pro Cys Glu Arg Ser Glu Thr Tyr
                595                 600                 605
gta aat ggc aag agg gtg tcc cag cct gtt cag ctg cgc tca gga aac      2052
Val Asn Gly Lys Arg Val Ser Gln Pro Val Gln Leu Arg Ser Gly Asn
            610                 615                 620
cgt atc atc atg ggt aaa aac cat gtt ttc cgc ttt aac cac ccg gaa      2100
Arg Ile Ile Met Gly Lys Asn His Val Phe Arg Phe Asn His Pro Glu
        625                 630                 635
caa gca cga gct gag cga gag aag act cct tct gct gag acc ccc tct      2148
Gln Ala Arg Ala Glu Arg Glu Lys Thr Pro Ser Ala Glu Thr Pro Ser
    640                 645                 650
gag cct gtg gac tgg aca ttt gcc cag agg gag ctt ctg gaa aaa caa      2196
Glu Pro Val Asp Trp Thr Phe Ala Gln Arg Glu Leu Leu Glu Lys Gln
655                 660                 665                 670
gga att gat atg aaa caa gag atg gag aaa agg cta cag gaa atg gag      2244
Gly Ile Asp Met Lys Gln Glu Met Glu Lys Arg Leu Gln Glu Met Glu
                675                 680                 685
atc tta tac aaa aag gag aag gaa gaa gca gat ctt ctt ttg gag cag      2292
Ile Leu Tyr Lys Lys Glu Lys Glu Glu Ala Asp Leu Leu Leu Glu Gln
            690                 695                 700
cag aga ctg gac tat gag agt aaa ttg cag gcc ttg cag aag cag gtt      2340
Gln Arg Leu Asp Tyr Glu Ser Lys Leu Gln Ala Leu Gln Lys Gln Val
        705                 710                 715
gaa acc cga tct ctg gct gca gaa aca act gaa gag gag gaa gaa gag      2388
Glu Thr Arg Ser Leu Ala Ala Glu Thr Thr Glu Glu Glu Glu Glu Glu
    720                 725                 730
gaa gaa gtt cct tgg aca cag cat gaa ttt gag ttg gcc caa tgg gcc      2436
Glu Glu Val Pro Trp Thr Gln His Glu Phe Glu Leu Ala Gln Trp Ala
735                 740                 745                 750
ttc cgg aaa tgg aag tct cat cag ttt act tca tta cgg gac tta ctc      2484
Phe Arg Lys Trp Lys Ser His Gln Phe Thr Ser Leu Arg Asp Leu Leu
                755                 760                 765
tgg ggc aat gcc gtg tac cta aag gag gcc aat gcc atc agt gtg gaa      2532
Trp Gly Asn Ala Val Tyr Leu Lys Glu Ala Asn Ala Ile Ser Val Glu
            770                 775                 780
ctg aaa aag aag gtg cag ttt cag ttt gtt ctg ctg act gac aca ctg      2580
Leu Lys Lys Lys Val Gln Phe Gln Phe Val Leu Leu Thr Asp Thr Leu
        785                 790                 795
tac tcc cct ttg cct cct gaa tta ctt ccc act gag atg gaa aaa act      2628
Tyr Ser Pro Leu Pro Pro Glu Leu Leu Pro Thr Glu Met Glu Lys Thr
    800                 805                 810
cat gag gac agg cct ttc cct cgc aca gtg gta gca gta gaa gtc cag      2676
His Glu Asp Arg Pro Phe Pro Arg Thr Val Val Ala Val Glu Val Gln
815                 820                 825                 830
gat ttg aag aat gga gca aca cac tat tgg tct ttg gag aaa ctc aag      2724
```

```
Asp Leu Lys Asn Gly Ala Thr His Tyr Trp Ser Leu Glu Lys Leu Lys
                835                 840                 845
cag agg ctg gat ttg atg cga gag atg tat gat agg gca ggg gag atg      2772
Gln Arg Leu Asp Leu Met Arg Glu Met Tyr Asp Arg Ala Gly Glu Met
            850                 855                 860
gcc tcc agt gcc caa gac gaa agc gaa acc act gtg act ggc agc gat      2820
Ala Ser Ser Ala Gln Asp Glu Ser Glu Thr Thr Val Thr Gly Ser Asp
        865                 870                 875
ccc ttc tat gat cgg ttc cac tgg ttc aaa ctt gtg ggg agc tcc ccc      2868
Pro Phe Tyr Asp Arg Phe His Trp Phe Lys Leu Val Gly Ser Ser Pro
    880                 885                 890
att ttc cac ggc tgt gtg aac gag cgc ctt gcc gac cgc aca ccc tcc      2916
Ile Phe His Gly Cys Val Asn Glu Arg Leu Ala Asp Arg Thr Pro Ser
895                 900                 905                 910
ccc act ttt tcc acg gcc gat tcc gac atc act gag ctg gct gac gag      2964
Pro Thr Phe Ser Thr Ala Asp Ser Asp Ile Thr Glu Leu Ala Asp Glu
                915                 920                 925
cag caa gat gag atg gag gat ttt gat gat gag gca ttc gtg gat gac      3012
Gln Gln Asp Glu Met Glu Asp Phe Asp Asp Glu Ala Phe Val Asp Asp
            930                 935                 940
gcc ggc tct gac gca ggg acg gag gag gga tca gat ctc ttc agt gac      3060
Ala Gly Ser Asp Ala Gly Thr Glu Glu Gly Ser Asp Leu Phe Ser Asp
        945                 950                 955
ggg cat gac ccg ttt tac gac cga tcc cct tgg ttc att tta gtg gga      3108
Gly His Asp Pro Phe Tyr Asp Arg Ser Pro Trp Phe Ile Leu Val Gly
    960                 965                 970
agg gca ttt gtt tac ctg agc aat ctg ctg tat ccc gtg ccc ctg atc      3156
Arg Ala Phe Val Tyr Leu Ser Asn Leu Leu Tyr Pro Val Pro Leu Ile
975                 980                 985                 990
cac agg gtg gcc atc gtc agt gag aaa ggt gaa gtg cgg gga ttt ctg      3204
His Arg Val Ala Ile Val Ser Glu Lys Gly Glu Val Arg Gly Phe Leu
                995                1000                1005
cgt gtg gct gta cag gcc atc gca gcg gat gaa gaa gct cct gat tat      3252
Arg Val Ala Val Gln Ala Ile Ala Ala Asp Glu Glu Ala Pro Asp Tyr
            1010                1015                1020
ggc tct gga att cga cag tca gga aca gct aaa ata tct ttt gat aat      3300
Gly Ser Gly Ile Arg Gln Ser Gly Thr Ala Lys Ile Ser Phe Asp Asn
        1025                1030                1035
gaa tac ttt aat cag agt gac ttt tcg tct gtt gca atg act cgt tct      3348
Glu Tyr Phe Asn Gln Ser Asp Phe Ser Ser Val Ala Met Thr Arg Ser
    1040                1045                1050
ggt ctg tcc ttg gag gag ttg agg att gtg gaa gga cag ggt cag agt      3396
Gly Leu Ser Leu Glu Glu Leu Arg Ile Val Glu Gly Gln Gly Gln Ser
1055                1060                1065                1070
tct gag gtc atc act cct cca gaa gaa atc agt cga att aat gac ttg      3444
Ser Glu Val Ile Thr Pro Pro Glu Glu Ile Ser Arg Ile Asn Asp Leu
                1075                1080                1085
gat ttg aag tca agc act ttg ctg gat ggt aag atg gta atg gaa ggg      3492
Asp Leu Lys Ser Ser Thr Leu Leu Asp Gly Lys Met Val Met Glu Gly
            1090                1095                1100
ttt tct gaa gag att ggc aac cac ctg aaa ctg ggc agt gcc ttc act      3540
Phe Ser Glu Glu Ile Gly Asn His Leu Lys Leu Gly Ser Ala Phe Thr
        1105                1110                1115
ttc cga gta aca gtg ttg cag gcc agt gga atc ctc cca gag tat gca      3588
Phe Arg Val Thr Val Leu Gln Ala Ser Gly Ile Leu Pro Glu Tyr Ala
    1120                1125                1130
gat atc ttc tgt cag ttc aac ttt ttg cat cgc cat gat gaa gca ttc      3636
Asp Ile Phe Cys Gln Phe Asn Phe Leu His Arg His Asp Glu Ala Phe
1135                1140                1145                1150
tcc acg gag ccc ctc aaa aac aat ggc aga gga agt ccc ctg gcc ttt      3684
Ser Thr Glu Pro Leu Lys Asn Asn Gly Arg Gly Ser Pro Leu Ala Phe
                1155                1160                1165
tat cat gtg cag aat att gca gtg gag atc act gaa tca ttt gtg gat      3732
```

```
Tyr His Val Gln Asn Ile Ala Val Glu Ile Thr Glu Ser Phe Val Asp
            1170                1175                1180
tac atc aaa acc aag cct att gta ttt gaa gtc ttt ggg cat tat cag      3780
Tyr Ile Lys Thr Lys Pro Ile Val Phe Glu Val Phe Gly His Tyr Gln
        1185                1190                1195
cag cac cca ctt cat ctg caa gga cag gag ctt aac agt ccg cct cag      3828
Gln His Pro Leu His Leu Gln Gly Gln Glu Leu Asn Ser Pro Pro Gln
    1200                1205                1210
ccg tgc cgc cga ttc ttc cct cca ccc atg cca ctg tcc aag cca gtt      3876
Pro Cys Arg Arg Phe Phe Pro Pro Pro Met Pro Leu Ser Lys Pro Val
1215                1220                1225                1230
cca gcc acc aag tta aac acg atg agc aaa acc agc ctt ggc cag agc      3924
Pro Ala Thr Lys Leu Asn Thr Met Ser Lys Thr Ser Leu Gly Gln Ser
                1235                1240                1245
atg agc aag tat gac ctc ctg gtt tgg ttt gag atc agt gaa ctg gag      3972
Met Ser Lys Tyr Asp Leu Leu Val Trp Phe Glu Ile Ser Glu Leu Glu
            1250                1255                1260
cct aca gga gag tat atc cca gct gtg gtt gac cac aca gca ggc ttg      4020
Pro Thr Gly Glu Tyr Ile Pro Ala Val Val Asp His Thr Ala Gly Leu
        1265                1270                1275
cct tgc cag ggg aca ttt ttg ctt cat cag ggc atc cag cga agg atc      4068
Pro Cys Gln Gly Thr Phe Leu Leu His Gln Gly Ile Gln Arg Arg Ile
    1280                1285                1290
aca gtg acc att atc cat gag aag ggg agc gag ctc cat tgg aaa gat      4116
Thr Val Thr Ile Ile His Glu Lys Gly Ser Glu Leu His Trp Lys Asp
1295                1300                1305                1310
gtt cgt gaa ctg gtg gta ggt cgt att cgg aat aag cct gag gtg gat      4164
Val Arg Glu Leu Val Val Gly Arg Ile Arg Asn Lys Pro Glu Val Asp
                1315                1320                1325
gaa gct gca gtt gat gcc atc ctc tcc cta aat att att tct gcc aag      4212
Glu Ala Ala Val Asp Ala Ile Leu Ser Leu Asn Ile Ile Ser Ala Lys
            1330                1335                1340
tac ctg aag tct tcc cac aac tct agc agg acc ttc tac cgc ttt gag      4260
Tyr Leu Lys Ser Ser His Asn Ser Ser Arg Thr Phe Tyr Arg Phe Glu
        1345                1350                1355
gct gtg tgg gat agc tct ctg cat aac tcc ctt ctt ctg aac cga gtg      4308
Ala Val Trp Asp Ser Ser Leu His Asn Ser Leu Leu Leu Asn Arg Val
    1360                1365                1370
aca ccc tat gga gaa aag atc tac atg acc ttg tcg gcc tac cta gag      4356
Thr Pro Tyr Gly Glu Lys Ile Tyr Met Thr Leu Ser Ala Tyr Leu Glu
1375                1380                1385                1390
ctg gat cat tgc atc cag ccg gct gtc atc acc aag gat gtg tgc atg      4404
Leu Asp His Cys Ile Gln Pro Ala Val Ile Thr Lys Asp Val Cys Met
                1395                1400                1405
gtc ttc tac tcc cga gat gcc aag atc tca cca cca cgc tct ctg cgt      4452
Val Phe Tyr Ser Arg Asp Ala Lys Ile Ser Pro Pro Arg Ser Leu Arg
            1410                1415                1420
agc ctc ttt ggc agc ggc tac tca aag tca cca gat tcg aat cga gtc      4500
Ser Leu Phe Gly Ser Gly Tyr Ser Lys Ser Pro Asp Ser Asn Arg Val
        1425                1430                1435
act ggc att tac gaa ctc agc tta tgc aaa atg tca gac aca ggt agt      4548
Thr Gly Ile Tyr Glu Leu Ser Leu Cys Lys Met Ser Asp Thr Gly Ser
    1440                1445                1450
cca ggt atg cag aga agg aga aga aaa atc tta gat acg tca gtg gca      4596
Pro Gly Met Gln Arg Arg Arg Arg Lys Ile Leu Asp Thr Ser Val Ala
1455                1460                1465                1470
tat gtg cgg gga gaa gag aac tta gca ggc tgg cgg ccc cgt gga gac      4644
Tyr Val Arg Gly Glu Glu Asn Leu Ala Gly Trp Arg Pro Arg Gly Asp
                1475                1480                1485
agc ctc atc ctt gag cac cag tgg gag ctg gag aag ctg gag ctc cta      4692
Ser Leu Ile Leu Glu His Gln Trp Glu Leu Glu Lys Leu Glu Leu Leu
            1490                1495                1500
cat gag gtg gaa aaa acc cgc cac ttt ttg ctg ctg cgt gag aga ctt      4740
```

```
His Glu Val Glu Lys Thr Arg His Phe Leu Leu Leu Arg Glu Arg Leu
        1505                1510                1515
ggt gac agc atc ccc aaa tcc ctg agc gac tcg tta tcc ccc agc ctc      4788
Gly Asp Ser Ile Pro Lys Ser Leu Ser Asp Ser Leu Ser Pro Ser Leu
    1520                1525                1530
agc agt ggg acc ctc agc acc tcc acc agt atc tcc tct cag atc tca      4836
Ser Ser Gly Thr Leu Ser Thr Ser Thr Ser Ile Ser Ser Gln Ile Ser
1535                1540                1545                1550
acc act acc ttt gaa agc gcc atc aca cct agc gag agc agt ggc tat      4884
Thr Thr Thr Phe Glu Ser Ala Ile Thr Pro Ser Glu Ser Ser Gly Tyr
                1555                1560                1565
gat tca gga gac atc gaa agc ctg gtg gac cga gag aaa gag ctg gct      4932
Asp Ser Gly Asp Ile Glu Ser Leu Val Asp Arg Glu Lys Glu Leu Ala
            1570                1575                1580
acc aag tgc ctg caa ctt ctc acc cac act ttc aac aga gaa ttc agc      4980
Thr Lys Cys Leu Gln Leu Leu Thr His Thr Phe Asn Arg Glu Phe Ser
        1585                1590                1595
cag gtg cac ggc agc gtc agt gac tgt aag ttg tct gat atc tct cca      5028
Gln Val His Gly Ser Val Ser Asp Cys Lys Leu Ser Asp Ile Ser Pro
    1600                1605                1610
att gga cgg gat ccc tct gag tcc agt ttc agc agt gcc acc ctc act      5076
Ile Gly Arg Asp Pro Ser Glu Ser Ser Phe Ser Ser Ala Thr Leu Thr
1615                1620                1625                1630
ccc tcc tcc acc tgt ccc tct ctg gta gac tct agg agc aac tct ctg      5124
Pro Ser Ser Thr Cys Pro Ser Leu Val Asp Ser Arg Ser Asn Ser Leu
                1635                1640                1645
gat cag aag acc cca gaa gcc aat tcc cgg gcc tct agt ccc tgc cca      5172
Asp Gln Lys Thr Pro Glu Ala Asn Ser Arg Ala Ser Ser Pro Cys Pro
            1650                1655                1660
gaa ttt gaa cag ttt cag att gtc cca gct gtg gaa aca cca tat ttg      5220
Glu Phe Glu Gln Phe Gln Ile Val Pro Ala Val Glu Thr Pro Tyr Leu
        1665                1670                1675
gcc cga gca gga aaa aac gaa ttt ctc aat ctt gtt cca gat att gaa      5268
Ala Arg Ala Gly Lys Asn Glu Phe Leu Asn Leu Val Pro Asp Ile Glu
    1680                1685                1690
gaa att aga cca agc tca gtg gtc tct aag aaa gga tac ctt cat ttc      5316
Glu Ile Arg Pro Ser Ser Val Val Ser Lys Lys Gly Tyr Leu His Phe
1695                1700                1705                1710
aag gag cct ctt tac agt aac tgg gct aaa cat ttt gtt gtc gtc cgt      5364
Lys Glu Pro Leu Tyr Ser Asn Trp Ala Lys His Phe Val Val Val Arg
                1715                1720                1725
cgg cct tat gtc ttc atc tat aac agt gac aaa gac cct gtg gag cgt      5412
Arg Pro Tyr Val Phe Ile Tyr Asn Ser Asp Lys Asp Pro Val Glu Arg
            1730                1735                1740
gga atc att aac ctg tcc aca gca cag gtg gag tac agt gag gac cag      5460
Gly Ile Ile Asn Leu Ser Thr Ala Gln Val Glu Tyr Ser Glu Asp Gln
        1745                1750                1755
cag gcc atg gtg aag aca cca aac acm ttt gct gtc tgc aca aag cac      5508
Gln Ala Met Val Lys Thr Pro Asn Thr Phe Ala Val Cys Thr Lys His
    1760                1765                1770
cgt ggg gtc ctt ttg cag gcc ctc aat gac aaa gac atg aac gac tgg      5556
Arg Gly Val Leu Leu Gln Ala Leu Asn Asp Lys Asp Met Asn Asp Trp
1775                1780                1785                1790
ttg tat gcc ttc aac cca ctt cta gct ggc aca ata cgg tca aag ctt      5604
Leu Tyr Ala Phe Asn Pro Leu Leu Ala Gly Thr Ile Arg Ser Lys Leu
                1795                1800                1805
tcc cgc aga tgc ccg agc cag tcg aaa tac taagtgactc tgccgagtgc        5654
Ser Arg Arg Cys Pro Ser Gln Ser Lys Tyr
            1810                1815
cctcactcgc cttcgagaga taaagaaagc gttacctctc atttctcttt gtgattcttg   5714
acggtgactc ttgtatgtaa tcctgtggct taactacttc tccctccttg tccagcactt   5774
ttctagctct cccgttcccc atctccattg ctctgtactc ttttcttttt tcttgtgctg   5834
agaatctcgt tagtagcatg tggcctaaca aaaggaaaaa atgtttttaa acacacacac   5894
```

acacacacac acacacacac acatacacan acaaaaacac aaaaactctg aggggatctg 5954
gtgaatctcc aaattattgt gggtgtactt tggcttcctt ttgtatgata ggtccccatc 6014
atgaccacct ctgatgtctg tgctgctgtc accaggcacc tttgtttttc aagacaacat 6074
actttttttt tcttttctct gtttgtgata tcactttaat ttttcttggg tggcttagag 6134
actaagggag gagacatctg gccttttag aacctgagag gaaaaaaaga gtctttttt 6194
cccctctgtc tctttttgcc atggctaatc cctgcatttc cattcaggga aaaggtggta 6254
gtgagcatag mactgcaaca gttatattct gagtcaaagt tggggctttt tacggcataa 6314
ttatggaatt tttatttact ggtagagagg agacgagagg ctttttcagt gggcctggga 6374
cagtggctgc tcttgacttt gtgtgaaggg aaatgccaag gatgcttctg gtggacttca 6434
ggggacccca gggtttggcc gtgggccgtg atggcagcag gcggtgggat gcttgtagct 6494
cctcacagca ggattcctgc ccactgtttt ttctctgttg ggagggaagc tcttttctag 6554
gagtgtctca gttctgcttt tggcattagt gatggtggtg gtacagttgg aattagtgcc 6614
atgtcataca caaatgttcc acaaggcggg agtgtttcac tttctggtga taaacttgat 6674
ggtcattgtt atgattaaga taatgccggg caggccgggc acagtggctc acgcctgtaa 6734
tccaagcact tggggaggcc gaggcgggca gatcacgaga tcaggagttc aagaccagcc 6794
tggccaatat gatgaaaccc cgtctctact aaaaatacaa aattagtcgg gtatggtggc 6854
acatgcctgt aattccagct gcttgggagc ctgaggcagg agaactgctt gaacccagga 6914
ggcagaggtt gcagtgagcc aagatcgcgc tattgcactc cagcctgggt gacagagcaa 6974
gactctgcct cagaaaaaaa aaaaaaaata ataatgctgg gtagtgacct tgtgattgtt 7034
acagctccct ttgatcaaag aaatatagct ttcaggcata aacctggaag tctccctctg 7094
aatccagcag ttgttttcat tgatgcttgt caggttgaag atgctttcag tgatgctctc 7154
tatactcata aataagcaaa atgtggcagg ctttgctttc tggatcccag gattaaaact 7214
aaccgtgacc actactccaa acaaaacaca atatgcctag gggcacggat gaacgtccag 7274
ggagcccggg ccccaggctt tgttgcgtgt tccctgctcc tctccatctg gtgtggaaac 7334
actgcccagg gagaaaggag gaagctcact gtggacagtc ttctttcctt ctgacagacc 7394
aggtcatctg gcttccgaga tcatcagaga agataagtct gtctctttca gctgccagta 7454
agttttccag gatgagaggg gaaaaagaaa gcctccagtg acttcagttg ctttgccagt 7514
tgtcttggga ttgttttaca ccatccttta cttcccttgc tcagacctct ctgtttcacc 7574
attgctcagg cattcaggaa agtatctgct cactcccact tggtgagtcc tcggccttga 7634
ggttgctgac tctcaggcgt taggcagctg gatgacttcc cgcttcatgc agcaaaggcc 7694
aggggcttgc gcgcctctgc agagttgttg ctagggagac ttgtgtcatc atccacaacc 7754
ttgtttctca cttcctggtt gggctcatct ctgaagaaca ggtctcccag cttcgctcct 7814
tatcactgca ttgtgaagag gaggaaaagt gaatcncgga gagagaaagg aaaggataga 7874
atcacaggct gcrtctgcac ctgaaaagtg acccgcggaa actctatggc ggatttttt 7934
tttaactttc ttcttcctgt taaaacatag gtcactaact gtgatgttat ttgttttcta 7994
agtggtatgt gagattttct aatgtagtta gaagtttcat tgtctgatgg acacaatatg 8054
cccttccggt tctattcaaa ccagcaggat ctgtcggtgc ttagagatgg ctgcctggac 8114
tggaatcaaa tctaatttca gggaaatgaa gatggaattt gaaggtcact tttaaaatta 8174
agtcattgat gctgctgtta cagagtgtga cagaggatcc atgtctgtga cacaggacgg 8234
tgggaagcct gagagagagt gaaattatgt gatacactga aatgactttt gtttttcttc 8294
taactcatac aaaactggtt tggaaagtct ttgctttgga agcgtcagac attagaacag 8354
gccaaactgg actgtctgtt catagcgtgc ctgaataaga aggcctctta gggagccaga 8414
gggagcagag tggtcgtgtc ctgcgtgctc ttcaccctct ggggcgcccc tgctgcggct 8474
ggcaggtgca gacagccttt gctggtcccc agcacgtcca gggtgggtgc tcccttgccc 8534
gacagaacca tccccactgt gaggctgtga gagatttgtg gcaggaactg tttatgaggc 8594
tctagttgtt gctgttgtgg cgggaaagtt aagaaacata gcccttaagg aaaccacctt 8654
tatgtatttt cttaaagcac gcctttaaat aagcaaaaac tttaaaaggc aggaaagaga 8714
attcttaggc aaattcagag aaataagtgc tagttaatac taatcacctc ctcctctgtc 8774
tctcatcctc ctttctccca tcaaagcaaa atatggcctc accaccagcc ccaaatcagt 8834
gctcagaccc tctctgtgtc tgtgtgccct cctgggagtc agtcagcgct caggccagga 8894
ctgtgcaggg ccagccagcc catgcgctag tcaggagcac aggcaagggg tgcttgtggc 8954
agtggccggg cacctgagcc ccagcttgtt gttaaacgtg ctgacggcaa ggggcaatgg 9014
agtgagtttc ccaactaaga aaccactatt atatattttt tcccttcagt cacatagact 9074
tcagacaact ctcctatttt ttatggattt ttcagctcat ttcagatgaa ggaactaagt 9134
cattgtgaac tgtctcttga gatctaaaaa caagatgact tttcctggca catattccaa 9194
agcaaagact ttgttgcctg ctgcttattg tctaatttac agggatattt aattttgtca 9254
ggtctatgta tatttatcca gctatactta cttgcacagt ggattggaga gaaaggattc 9314
tccagtgtgc acactcatcg gtactctttc tgcatttccc tcgtgctgtg tcccgctcgg 9374
gttccaatgg acagtatcag ggcttgtttg acttaggtct ttcagttttc ctttcggttc 9434
ccttttaaaa atgtgattgt taacctgcct cttgaaagat tcaaccgggt gtggtggctc 9494
acgcctgtaa tcccagcaca ttgggaggcc aaggcaggtg gatcacctga ggtcaggagt 9554
ttgagaccag cctggccaac atggtgaaac cccgtctcta ctaaaaatac aaaaagtagc 9614
caggcgtggt ggcgtgtgcc tgtagtccca gctacttggg aggctgaggc aggagaatag 9674

cttgaacctg ggaggtggag gctgcagtga gttgagactg caccattgca ctccagcctg 9734
ggtgacaaag caggactccg tctcaaaaaa aaagaaaaga ttcatgatgc tgctgctccc 9794
agaaggtttg ctggatgtgt ttacatagga ctctaacttg tgtgcactac agttgttcac 9854
cagggccagt gattcacccc agtgtgtggc cagaccatga ctgtgtagca ggaatgtttt 9914
aatttgtgct tccttagtaa attgaaatat cagctgagag attatttgct gctgttattc 9974
aaaaggccat ttatgaagtt agtatttgag ccccataaga tctttaaaaa gcctccaatc 10034
atttaaagga agaaatcaga gttgctataa aattcagtaa aaagctcata gccaaaacgg 10094
ctgtgctcag atggaaagtc tgagctgagg ttggtctctt gccaaaccgt ggctgttgtg 10154
tgttgttctt catgtcttcg agttcatttt ttttcattct gcctattctg gcatcagctc 10214
acttgaggag tccctcagcc ttcttgtatt taaggcatcg tcttagactt tgtggctcta 10274
aagtacctgt ctgttgagat ttcaagtctc ttgtcaccat cctcacacat gacaacaaaa 10334
cccataatgc ataagtggcc tttttgaacc aagactttgc aaactgatct ctcccccgtg 10394
aaggagttga gcacattagc aacaatgtac attaattttg gattttcatt ttcatgtttt 10454
attttgtaaa tattatctga tgtttggagc ttgagtatac agactgtaaa tatagttctt 10514
gtatttgtac taattctgat tcttttgctg tatagcctta gatgtgcaat gcagacacta 10574
tctaactgtg tgtggtaacc ttgcgtcacg gagctgttag tgaacgaggt aaaaataata 10634
aaggtacagc cagtgcatca aaaaaaaaaa aaaaaaaaaa aaaaaaaa 10682

<210> 4
<211> 1816
<212> PRT
<213> Homo sapiens

<400> 4
Met Ser Gly Ala Ser Val Lys Val Ala Val Arg Val Arg Pro Phe Asn
1 5 10 15
Ser Arg Glu Thr Ser Lys Glu Ser Lys Cys Ile Ile Gln Met Gln Gly
20 25 30
Asn Ser Thr Ser Ile Ile Asn Pro Lys Asn Pro Lys Glu Ala Pro Lys
35 40 45
Ser Phe Ser Phe Asp Tyr Ser Tyr Trp Ser His Thr Ser Pro Glu Asp
50 55 60
Pro Cys Phe Ala Ser Gln Asn Arg Val Tyr Asn Asp Ile Gly Lys Glu
65 70 75 80
Met Leu Leu His Ala Phe Glu Gly Tyr Asn Val Cys Ile Phe Ala Tyr
85 90 95
Gly Gln Thr Gly Ala Gly Lys Ser Tyr Thr Met Met Gly Lys Gln Glu
100 105 110
Glu Ser Gln Ala Gly Ile Ile Pro Gln Leu Cys Glu Glu Leu Phe Glu
115 120 125
Lys Ile Asn Asp Asn Cys Asn Glu Glu Met Ser Tyr Ser Val Glu Val
130 135 140
Ser Tyr Met Glu Ile Tyr Cys Glu Arg Val Arg Asp Leu Leu Asn Pro
145 150 155 160
Lys Asn Lys Gly Asn Leu Arg Val Arg Glu His Pro Leu Leu Gly Pro
165 170 175
Tyr Val Glu Asp Leu Ser Lys Leu Ala Val Thr Ser Tyr Thr Asp Ile
180 185 190
Ala Asp Leu Met Asp Ala Gly Asn Lys Ala Arg Thr Val Ala Ala Thr
195 200 205
Asn Met Asn Glu Thr Ser Ser Arg Ser His Ala Val Phe Thr Ile Val
210 215 220
Phe Thr Gln Lys Lys His Asp Asn Glu Thr Asn Leu Ser Thr Glu Lys
225 230 235 240
Val Ser Lys Ile Ser Leu Val Asp Leu Ala Gly Ser Glu Arg Ala Asp
245 250 255
Ser Thr Gly Ala Lys Gly Thr Arg Leu Lys Glu Gly Ala Asn Ile Asn
260 265 270
Lys Ser Leu Thr Thr Leu Gly Lys Val Ile Ser Ala Leu Ala Glu Val
275 280 285
Asp Asn Cys Thr Ser Lys Ser Lys Lys Lys Lys Lys Thr Asp Phe Ile
290 295 300
Pro Tyr Arg Asp Ser Val Leu Thr Trp Leu Leu Arg Glu Asn Leu Gly

305 310 315 320
Gly Asn Ser Arg Thr Ala Met Val Ala Ala Leu Ser Pro Ala Asp Ile
325 330 335
Asn Tyr Asp Glu Thr Leu S r Thr Leu Arg Tyr Ala Asp Arg Ala Lys
340 345 350
Gln Ile Lys Cys Asn Ala Val Ile Asn Glu Asp Pro Asn Ala Lys Leu
355 360 365
Val Arg Glu Leu Lys Glu Glu Val Thr Arg Leu Lys Asp Leu Leu Arg
370 375 380
Ala Gln Gly Leu Gly Asp Ile Ile Asp Ile Asp Pro Leu Ile Asp Asp
385 390 395 400
Tyr Ser Gly Ser Gly Ser Lys Tyr Leu Lys Asp Phe Gln Asn Asn Lys
405 410 415
His Arg Tyr Leu Leu Ala Ser Glu Asn Gln Arg Pro Gly His Phe Ser
420 425 430
Thr Ala Ser Met Gly Ser Leu Thr Ser Ser Pro Ser Ser Cys Ser Leu
435 440 445
Ser Ser Gln Val Gly Leu Thr Ser Val Thr Ser Ile Gln Glu Arg Ile
450 455 460
Met Ser Thr Pro Gly Gly Glu Glu Ala Ile Glu Arg Leu Lys Glu Ser
465 470 475 480
Glu Lys Ile Ile Ala Glu Leu Asn Glu Thr Trp Glu Glu Lys Leu Arg
485 490 495
Lys Thr Glu Ala Ile Arg Met Glu Arg Glu Ala Leu Leu Ala Glu Met
500 505 510
Gly Val Ala Ile Arg Glu Asp Gly Gly Thr Leu Gly Val Phe Ser Pro
515 520 525
Lys Lys Thr Pro His Leu Val Asn Leu Asn Glu Asp Pro Leu Met Ser
530 535 540
Glu Cys Leu Leu Tyr Tyr Ile Lys Asp Gly Ile Thr Arg Val Gly Gln
545 550 555 560
Ala Asp Ala Glu Arg Arg Gln Asp Ile Val Leu Ser Gly Ala His Ile
565 570 575
Lys Glu Glu His Cys Ile Phe Arg Ser Glu Arg Ser Asn Ser Gly Glu
580 585 590
Val Ile Val Thr Leu Glu Pro Cys Glu Arg Ser Glu Thr Tyr Val Asn
595 600 605
Gly Lys Arg Val Ser Gln Pro Val Gln Leu Arg Ser Gly Asn Arg Ile
610 615 620
Ile Met Gly Lys Asn His Val Phe Arg Phe Asn His Pro Glu Gln Ala
625 630 635 640
Arg Ala Glu Arg Glu Lys Thr Pro Ser Ala Glu Thr Pro Ser Glu Pro
645 650 655
Val Asp Trp Thr Phe Ala Gln Arg Glu Leu Leu Glu Lys Gln Gly Ile
660 665 670
Asp Met Lys Gln Glu Met Glu Lys Arg Leu Gln Glu Met Glu Ile Leu
675 680 685
Tyr Lys Lys Glu Lys Glu Glu Ala Asp Leu Leu Leu Glu Gln Gln Arg
690 695 700
Leu Asp Tyr Glu Ser Lys Leu Gln Ala Leu Gln Lys Gln Val Glu Thr
705 710 715 720
Arg Ser Leu Ala Ala Glu Thr Thr Glu Glu Glu Glu Glu Glu Glu Glu
725 730 735
Val Pro Trp Thr Gln His Glu Phe Glu Leu Ala Gln Trp Ala Phe Arg
740 745 750
Lys Trp Lys Ser His Gln Phe Thr Ser Leu Arg Asp Leu Leu Trp Gly
755 760 765
Asn Ala Val Tyr Leu Lys Glu Ala Asn Ala Ile Ser Val Glu Leu Lys
770 775 780
Lys Lys Val Gln Phe Gln Phe Val Leu Leu Thr Asp Thr Leu Tyr Ser
785 790 795 800
Pro Leu Pro Pro Glu Leu Leu Pro Thr Glu Met Glu Lys Thr His Glu
805 810 815

Asp Arg Pro Phe Pro Arg Thr Val Val Ala Val Glu Val Gln Asp Leu
820 825 830
Lys Asn Gly Ala Thr His Tyr Trp Ser Leu Glu Lys Leu Lys Gln Arg
835 840 845
Leu Asp Leu Met Arg Glu Met Tyr Asp Arg Ala Gly Glu Met Ala Ser
850 855 860
Ser Ala Gln Asp Glu Ser Glu Thr Thr Val Thr Gly Ser Asp Pro Phe
865 870 875 880
Tyr Asp Arg Phe His Trp Phe Lys Leu Val Gly Ser Ser Pro Ile Phe
885 890 895
His Gly Cys Val Asn Glu Arg Leu Ala Asp Arg Thr Pro Ser Pro Thr
900 905 910
Phe Ser Thr Ala Asp Ser Asp Ile Thr Glu Leu Ala Asp Glu Gln Gln
915 920 925
Asp Glu Met Glu Asp Phe Asp Asp Glu Ala Phe Val Asp Asp Ala Gly
930 935 940
Ser Asp Ala Gly Thr Glu Glu Gly Ser Asp Leu Phe Ser Asp Gly His
945 950 955 960
Asp Pro Phe Tyr Asp Arg Ser Pro Trp Phe Ile Leu Val Gly Arg Ala
965 970 975
Phe Val Tyr Leu Ser Asn Leu Leu Tyr Pro Val Pro Leu Ile His Arg
980 985 990
Val Ala Ile Val Ser Glu Lys Gly Glu Val Arg Gly Phe Leu Arg Val
995 1000 1005
Ala Val Gln Ala Ile Ala Ala Asp Glu Glu Ala Pro Asp Tyr Gly Ser
1010 1015 1020
Gly Ile Arg Gln Ser Gly Thr Ala Lys Ile Ser Phe Asp Asn Glu Tyr
1025 1030 1035 1040
Phe Asn Gln Ser Asp Phe Ser Ser Val Ala Met Thr Arg Ser Gly Leu
1045 1050 1055
Ser Leu Glu Glu Leu Arg Ile Val Glu Gly Gln Gly Gln Ser Ser Glu
1060 1065 1070
Val Ile Thr Pro Pro Glu Glu Ile Ser Arg Ile Asn Asp Leu Asp Leu
1075 1080 1085
Lys Ser Ser Thr Leu Leu Asp Gly Lys Met Val Met Glu Gly Phe Ser
1090 1095 1100
Glu Glu Ile Gly Asn His Leu Lys Leu Gly Ser Ala Phe Thr Phe Arg
1105 1110 1115 1120
Val Thr Val Leu Gln Ala Ser Gly Ile Leu Pro Glu Tyr Ala Asp Ile
1125 1130 1135
Phe Cys Gln Phe Asn Phe Leu His Arg His Asp Glu Ala Phe Ser Thr
1140 1145 1150
Glu Pro Leu Lys Asn Asn Gly Arg Gly Ser Pro Leu Ala Phe Tyr His
1155 1160 1165
Val Gln Asn Ile Ala Val Glu Ile Thr Glu Ser Phe Val Asp Tyr Ile
1170 1175 1180
Lys Thr Lys Pro Ile Val Phe Glu Val Phe Gly His Tyr Gln Gln His
1185 1190 1195 1200
Pro Leu His Leu Gln Gly Gln Glu Leu Asn Ser Pro Pro Gln Pro Cys
1205 1210 1215
Arg Arg Phe Phe Pro Pro Pro Met Pro Leu Ser Lys Pro Val Pro Ala
1220 1225 1230
Thr Lys Leu Asn Thr Met Ser Lys Thr Ser Leu Gly Gln Ser Met Ser
1235 1240 1245
Lys Tyr Asp Leu Leu Val Trp Phe Glu Ile Ser Glu Leu Glu Pro Thr
1250 1255 1260
Gly Glu Tyr Ile Pro Ala Val Val Asp His Thr Ala Gly Leu Pro Cys
1265 1270 1275 1280
Gln Gly Thr Phe Leu Leu His Gln Gly Ile Gln Arg Arg Ile Thr Val
1285 1290 1295
Thr Ile Ile His Glu Lys Gly Ser Glu Leu His Trp Lys Asp Val Arg
1300 1305 1310
Glu Leu Val Val Gly Arg Ile Arg Asn Lys Pro Glu Val Asp Glu Ala

1315 1320 1325
Ala Val Asp Ala Ile Leu Ser Leu Asn Ile Ile Ser Ala Lys Tyr Leu
1330 1335 1340
Lys Ser Ser His Asn Ser Ser Arg Thr Phe Tyr Arg Phe Glu Ala Val
1345 1350 1355 1360
Trp Asp Ser Ser Leu His Asn Ser Leu Leu Leu Asn Arg Val Thr Pro
1365 1370 1375
Tyr Gly Glu Lys Ile Tyr Met Thr Leu Ser Ala Tyr Leu Glu Leu Asp
1380 1385 1390
His Cys Ile Gln Pro Ala Val Ile Thr Lys Asp Val Cys Met Val Phe
1395 1400 1405
Tyr Ser Arg Asp Ala Lys Ile Ser Pro Pro Arg Ser Leu Arg Ser Leu
1410 1415 1420
Phe Gly Ser Gly Tyr Ser Lys Ser Pro Asp Ser Asn Arg Val Thr Gly
1425 1430 1435 1440
Ile Tyr Glu Leu Ser Leu Cys Lys Met Ser Asp Thr Gly Ser Pro Gly
1445 1450 1455
Met Gln Arg Arg Arg Arg Lys Ile Leu Asp Thr Ser Val Ala Tyr Val
1460 1465 1470
Arg Gly Glu Glu Asn Leu Ala Gly Trp Arg Pro Arg Gly Asp Ser Leu
1475 1480 1485
Ile Leu Glu His Gln Trp Glu Leu Glu Lys Leu Glu Leu Leu His Glu
1490 1495 1500
Val Glu Lys Thr Arg His Phe Leu Leu Leu Arg Glu Arg Leu Gly Asp
1505 1510 1515 1520
Ser Ile Pro Lys Ser Leu Ser Asp Ser Leu Ser Pro Ser Leu Ser Ser
1525 1530 1535
Gly Thr Leu Ser Thr Ser Thr Ser Ile Ser Ser Gln Ile Ser Thr Thr
1540 1545 1550
Thr Phe Glu Ser Ala Ile Thr Pro Ser Glu Ser Ser Gly Tyr Asp Ser
1555 1560 1565
Gly Asp Ile Glu Ser Leu Val Asp Arg Glu Lys Glu Leu Ala Thr Lys
1570 1575 1580
Cys Leu Gln Leu Leu Thr His Thr Phe Asn Arg Glu Phe Ser Gln Val
1585 1590 1595 1600
His Gly Ser Val Ser Asp Cys Lys Leu Ser Asp Ile Ser Pro Ile Gly
1605 1610 1615
Arg Asp Pro Ser Glu Ser Ser Phe Ser Ser Ala Thr Leu Thr Pro Ser
1620 1625 1630
Ser Thr Cys Pro Ser Leu Val Asp Ser Arg Ser Asn Ser Leu Asp Gln
1635 1640 1645
Lys Thr Pro Glu Ala Asn Ser Arg Ala Ser Ser Pro Cys Pro Glu Phe
1650 1655 1660
Glu Gln Phe Gln Ile Val Pro Ala Val Glu Thr Pro Tyr Leu Ala Arg
1665 1670 1675 1680
Ala Gly Lys Asn Glu Phe Leu Asn Leu Val Pro Asp Ile Glu Glu Ile
1685 1690 1695
Arg Pro Ser Ser Val Val Ser Lys Lys Gly Tyr Leu His Phe Lys Glu
1700 1705 1710
Pro Leu Tyr Ser Asn Trp Ala Lys His Phe Val Val Val Arg Arg Pro
1715 1720 1725
Tyr Val Phe Ile Tyr Asn Ser Asp Lys Asp Pro Val Glu Arg Gly Ile
1730 1735 1740
Ile Asn Leu Ser Thr Ala Gln Val Glu Tyr Ser Glu Asp Gln Gln Ala
1745 1750 1755 1760
Met Val Lys Thr Pro Asn Thr Phe Ala Val Cys Thr Lys His Arg Gly
1765 1770 1775
Val L u Leu Gln Ala Leu Asn Asp Lys Asp Met Asn Asp Trp Leu Tyr
1780 1785 1790
Ala Phe Asn Pro Leu Leu Ala Gly Thr Ile Arg Ser Lys Leu Ser Arg
1795 1800 1805
Arg Cys Pro Ser Gln Ser Lys Tyr
1810 1815

<210> 5
<211> 357
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> 178
<223> 10-265-178 : polymorphic base A or G

<220>
<221> misc_binding
<222> 166..190
<223> 10-265-178.probe

<220>
<221> primer_bind
<222> 159..177
<223> 10-265-178.mis

<220>
<221> primer_bind
<222> 179..197
<223> 10-265-178.mis complement

<220>
<221> primer_bind
<222> 1..18
<223> 10-265.pu

<220>
<221> primer_bind
<222> 338..357
<223> 10-265.rp complement

<400> 5
ccaaatgtgg aaggaccgga cccctgggtt gcagcgcgtc gagcggtgct gactctttcc 60
tttgttctgt ttctgcctct ctagagctga catcgcgctg atcggattgg ccgtcatggg 120
ccagaactta attctgaaca tgaatgacca cggctttgtg gtaagcggcg tgggcgcrtt 180
gtcttctctc tggttcccgg gcgctttagc cgaggccggc gataggtttg ggagcttacg 240
ggtctcctgg ccgtgctttg ctaatgtgct ctgttgctgc tcgtggcatt tttgtatgga 300
aaggagaagc accctgtagg cgtgggcggc cgatcccgaa cttagtcctg cggagtg 357

<210> 6
<211> 420
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> 203
<223> 10-266-203 : polymorphic base C or T

<220>
<221> misc_binding
<222> 191..215
<223> 10-266-203.probe

<220>
<221> primer_bind
<222> 184..202
<223> 10-266-203.mis

<220>
<221> primer_bind
<222> 204..222
<223> 10-266-203.mis complement

<220>
<221> primer_bind
<222> 1..20
<223> 10-266.pu

<220>
<221> primer_bind
<222> 401..420
<223> 10-266.rp complement

<400> 6
ggtgacataa ctttacagtg aagctcccac agctggggga agaaaaggca aaggcaggtc 60
tgacctcccc agaacttggg ttgaatggaa aggtcattgt tactttggcc acagtctgaa 120
agtcttgtgt gtcctggatc tcctactcag gacttttgtc cttctaggtc tgtgctttta 180
ataggactgt ctccaaagtt gaygatttct tggccaatga ggcaaaggga accaaagtgg 240
tgggtgccca gtccctgaaa gagatggtct ccaagctgaa gaagccccgg cggatcatcc 300
tcctggtgaa ggctgggcaa gctgtggatg atttcatcga gaaattggtg aggccagctg 360
tgctctcagc tgctaccacg atagcagctg tttttggttt cttcctttag ttctccttct 420

<210> 7
<211> 465
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> 118
<223> 12-592-118 : polymorphic base A or T

<220>
<221> misc_binding
<222> 106..130
<223> 12-592-118.probe

<220>
<221> primer_bind
<222> 99..117
<223> 12-592-118.mis

<220>
<221> primer_bind
<222> 119..137
<223> 12-592-118.mis complement

<220>
<221> primer_bind
<222> 1..19
<223> 12-592.pu

<220>
<221> primer_bind
<222> 448..465
<223> 12-592.rp complement

<400> 7
gagttcagtt tgagttcagc tacgagaatt atttagttga gtgagtgtca gagctgggat 60
tttcaaatct gcccttactg gtatgttgct ttacacaccc tcttacgtaa atcaaacwag 120

gattccaccc tgggaggttt gcatagaggg ctgttcttgt agaacttgtg ctcatgcttt 180
gatttgggat ttggggagtt agggcaagcc agaaagtttt tctggtggat aataatgtgg 240
gttgactttc ttaagcattt taagccaagc acttgagttt ctaacaacta aaaagctaag 300
tcagcctgac acagctctag cgcgccctgg cttgattctg ttcatcccca gggggagact 360
tgcctttgtt ccagtcctgc tctcccaagc cagcttactg tagttttcca gcaattctga 420
gaagcagtat tttttactgc tgattagacc ttaacatgga aatgg 465

<210> 8
<211> 449
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> 420
<223> 12-783-421 : polymorphic base C or T

<220>
<221> misc_binding
<222> 408..432
<223> 12-783-421.probe

<220>
<221> primer_bind
<222> 401..419
<223> 12-783-421.mis

<220>
<221> primer_bind
<222> 421..439
<223> 12-783-421.mis complement

<220>
<221> primer_bind
<222> 1..19
<223> 12-783.pu

<220>
<221> primer_bind
<222> 429..449
<223> 12-783.rp complement

<220>
<221> allele
<222> 72
<223> 12-783-73 : polymorphic base G or C

<220>
<221> misc_binding
<222> 60..84
<223> 12-783-73.probe

<220>
<221> primer_bind
<222> 53..71
<223> 12-783-73.mis

<220>
<221> primer_bind
<222> 73..91
<223> 12-783-73.mis complement

<400> 8

atgactgtga gaacgaaggg gtaaactggg acaagatccg gtgattgcag aaggcttctg 60
aggaggcgac astgcacaga gacctaagtg aagtgaggag cccttgcctc tgaggagaca 120
acacttctgg tggccagaaa ttcagtataa ccaggttcag aacaaacaca actgactctg 180
ggttagcata aaactcaacc agcaggagca gaagtcccca ggcagacgcg agcagagcct 240
gctggcaacc gtgagcattg gtcagcgtgg acattggaca aggggcttcc ttgctcagcc 300
ctcattcctc taacatggtt ctctcctgtg ttctgcatgt aggcctttga ggattggaat 360
aagacagagc tagactcatt cctgattgaa atcacagcca atattctcaa gttccaagay 420
accgatggca aacacctgct gccaaagat 449

<210> 9
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerPU

<400> 9
tgtaaaacga cggccagt 18

<210> 10
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerRP

<400> 10
caggaaacag ctatgacc 18

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 00/00562

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 A01K67/027 C07K16/18 C12Q1/68
G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A01K C12Q G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
STRAND, BIOSIS, EMBASE, CAB Data, EPO-Internal, PAJ, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | Y. OKADA ET AL.: "The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors." CELL, vol. 81, February 1906 (1906-02), pages 769-780, XP000929277 the whole document<br>---<br>-/-- | 1-67 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 4 September 2000 | 08/09/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Hix, R |

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

national Application No
PCT/IB 00/00562

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | M. NANGAKU ET AL.: "KIF1B, a novel microtubule plus end-directed monomeric motor protein for transport of mitochondria." CELL, vol. 79, 30 December 1994 (1994-12-30), pages 1209-1220, XP000915469 cited in the application the whole document | 1-67 |
| Y | R.A. FURLONG ET AL.: "Characterization of a kinesin-related gene ATSV, within the tuberous sclerosis locus (TSC1) candidate region on chromosome 9Q34." GENOMICS, vol. 33, no. 3, 1 May 1996 (1996-05-01), pages 421-429, XP000925509 the whole document | 1-67 |
| Y | AIZAWA H ET AL: "Kinesin family in murine central nervous system" JOURNAL OF CELL BIOLOGY,US,ROCKEFELLER UNIVERSITY PRESS, NEW YORK, US, vol. 119, 1 December 1992 (1992-12-01), pages 1287-1296, XP002080692 ISSN: 0021-9525 the whole document | 1-67 |
| P,X | T-W. L. GONG ET AL.: "A novel mouse kinesin of the UNC-104/KIF1 subfamily encoded by the Kif1b gene" GENE, vol. 239, 1999, pages 117-127, XP000925542 the whole document | 17-22, 46,47 |
| A | Y. YONEKAWA ET AL.: "Defect in synaptic vesicle precursor transport and neuronal cell death in KIF1A motor protein-deficient mice." JOURNAL OF CELL BIOLOGY, vol. 141, no. 2, 20 April 1998 (1998-04-20), pages 431-441, XP000925196 cited in the application the whole document | |
| P,A | WO 00 18919 A (INCYTE PHARMA INC ;PATTERSON CHANDRA (US); CORLEY NEIL C (US); GUE) 6 April 2000 (2000-04-06) the whole document | |

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/IB 00/00562

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 0018919 A | 06-04-2000 | US 6013454 A | 11-01-2000 |
| | | AU 6403599 A | 17-04-2000 |